

VISIT...

LANZAROTE
Caliente.COM

Neural Network Solutions for Trading in Financial Markets

DIRK EMMA BAESTAENS
WILLEM MAX VAN DEN BERGH
DOUGLAS WOOD

PITMAN PUBLISHING

НЕЙРОННЫЕ СЕТИ И ФИНАНСОВЫЕ РЫНКИ

Принятие решений в торговых операциях

Д.-Э. БЭСТЕНС, В.-М. ВАН ДЕН БЕРГ, Д. ВУД

Перевод с английского *С. В. Курочкина*
под редакцией *А. П. Коваленко и Е. С. Пастухова*

ТВП
Научное
издательство
Москва

УДК 331:33+336.2+336.76/77+517.2+681.14

Научное издательство ТВП
ул. Губкина 8
117966 Москва ГСП-1
Россия

© Longman Group Ltd 1994. Assigned to Pearson Professional Ltd 1995.

This translation of *Baestaens, Van den Bergh & Wood: Neural Network Solutions for Trading in Financial Markets* is published by arrangement with Pitman Publishing, a division of Pearson Professional Limited, London

© Перевод на русский язык. Научное изд-во «ТВП» 1997.

Все права защищены. Никакая часть этой книги не может быть воспроизведена, сохранена в запоминающих системах или передана в любой форме и любым способом — электронным, механическим, фотокопировальным, записывающим или иным — без предварительного письменного разрешения владельца прав. За информацией обращаться по адресу: 117966 Москва ГСП-1, ул. Губкина 8, ТВП.

Бэстэнс Д.-Э., ван ден Берг В.-М., Вуд Д. Нейронные сети и финансовые рынки: принятие решений в торговых операциях. — Москва: ТВП, 1997. — xx, 236 с.

ISBN 5-85484-028-6

Нейронно-сетевая методология, пока мало представленная в российской профессиональной научно-технической литературе, находит все новые успешные применения в практике управления и принятия решений, в том числе — в финансовой и торговой сферах. Лежащая в ее основе теория нелинейных адаптивных систем доказала свою полезность при выработке прогнозов в целом ряде отраслей экономики и финансов.

Книга знакомит со способами применения методологии нейронных сетей для решения задач анализа и прогноза в таких актуальных для современной российской экономики вопросах, как кризисные явления на рынках капитала, налоговые поступления, динамика цен производных финансовых инструментов и индексов курсов акций, эффективность диверсификации портфельных капиталовложений, риск предоставления кредитов или банкротство корпораций и банков. Постоянные сравнения с иными применяемыми способами анализа и прогноза (например, статистическими способами анализа временных рядов и классификации или способами технического анализа) помогают читателю точнее определить роль и место нейронно-сетевых методов в областях, представляющих для него практический интерес.

Данное издание адресовано, в первую очередь, финансовым директорам, управляющим и аналитикам финансовых организаций, специалистам по количественному анализу и системным экспертам, а также студентам и аспирантам соответствующих специальностей.

Ил. 51. Библиогр. 296 назв.

Б 0607000000-28
Ю54 (03)-97

ISBN 5-85484-028-6

Выпускающие редакторы: Л. И. Герасимова и В. И. Хохлов.

Титульные редакторы: А. П. Коваленко и Е. С. Пастухов.

Научные редакторы: Л. В. Мищенко и В. И. Хохлов.

Композиция: Е. В. Герасимова и Н. В. Третьяков.

Худ. оформление: Е. В. Герасимова и А. Я. Галкин.

Набор: ООО «Технополис» и ООО «Скайрум».

Лицензия ЛР-061510 от 08.09.97

Подписано к печати 17.12.97. Формат 70×100/16. Бумага Люминофсет.

Печать офсетная. Гарнитура Миньон. Усл. печ. л. 21,2.

Усл. кр.-отт. 21,2 тыс. Уч.-изд. л. 21,3. Бум. л. 8,0.

Тираж 1000 экз. Зак. 5928. Цена договорная.

АО «Типография «Новости».

107005 Москва, ул. Ф.Энгельса, 46

Оглавление

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ: “POVERKH BAR'EROV”	vii
ОБ АВТОРАХ	x
ПРЕДИСЛОВИЕ	xii
ВВЕДЕНИЕ	xv
Глава 1. НЕЙРОННО-СЕТЕВЫЕ МЕТОДЫ	1
Введение в методы нейронных сетей	1
Устройство нейронных сетей	3
Обучение	7
Обобщающие правила	15
Примечания	25
Глава 2. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ КЛАССИФИКАЦИИ И АНАЛИЗА ВРЕМЕННЫХ РЯДОВ	26
Нейронные сети в задачах классификации	26
Применение нейронных сетей в анализе временных рядов	38
Сравнительная оценка производительности нейронных сетей	50
Программное обеспечение	53
Примечания	55
Глава 3. БАНКРОТСТВА, ПАНИКИ И БЕЗУМИЯ	56
Теория хаоса и рынки капитала	56
Банкротства, паники и безумия	58
Можно ли предсказывать закономерности во временном ряде цен	62
Несколько нейронно-сетевых экспериментов с логистическими временными рядами	63
Сетевая оценка в двумерной задаче (отображение Хенона)	68
Упрощенный вариант модели Хенона	73
Некоторые итоговые замечания	76
Примечания	77
Глава 4. ПРОГНОЗИРОВАНИЕ ДЕНЕЖНЫХ ПОТОКОВ. НАЛОГОВЫЕ ПОСТУПЛЕНИЯ	78
Голландская нормативная база	78
Традиционные методы оценки	79
Выбор переменных	80
Нейронно-сетевая модель	83
Вклад каждой из переменных по отдельности	88
Выводы	95
Примечания	95
Глава 5. ВРЕМЕННЫЕ РЯДЫ В ЗАДАЧАХ РАСЧЕТА ЦЕН ОПЦИОНОВ ЕВРОПЕЙСКОГО ТИПА	96
Постановка задачи	96
Теоретические основы	97
Эндогенные и экзогенные переменные	101
Предварительная обработка данных и подготовительные тесты	111
Результаты работы сети	114
Обсуждение	116

Глава 6. ОЦЕНКА ИНДЕКСОВ КУРСОВ АКЦИЙ

Влияние экономических факторов и построение моделей	118
Линейная модель АРТ	121
Многослойная схема с обратным распространением ошибки	123
Сравнение индивидуального и систематического вклада переменных	128
Выводы	135

Глава 7. УПРАВЛЕНИЕ МЕЖДУНАРОДНЫМ ПОРТФЕЛЕМ

Интернационализация портфельных инвестиций	138
Способы оценки результатов	140
Формирование портфеля: экспертное мнение	142
Спецификация модели	143
Предварительная обработка	144
Обучение	145
Результаты	146
Анализ результатов	147
Выводы	149

Глава 8. ОЦЕНКА КРЕДИТНОГО РИСКА НА ОСНОВАНИИ ДАННЫХ НЕФИНАНСОВОГО ХАРАКТЕРА

Модели предсказания банкротств	152
Предоставление займов малым и средним предприятиям (опыт Польши)	154
Описание базы данных	156
MDA как точка отсчета	161
Нейронно-сетевые модели	162
Обсуждение результатов	167
Опыт оценки кредитного риска в Голландском инвестиционном банке	168
Описание базы данных Голландского инвестиционного банка	169
Две точки отсчета: MDA и kNN	172
Результаты классификации с помощью нейронных сетей	173
Обсуждение	177
Приложения	179

Глава 9. ПРОГНОЗИРОВАНИЕ БАНКРОТСТВА КОРПОРАЦИЙ

Возможности нейронных сетей в задаче прогнозирования банкротства корпораций	183
Оценка качества моделей	185
Эксперимент	187
Разработка модели	189
Сравнение результатов	190
Использование в нейронной сети пониженных разделяющих уровней	191

Глава 10. ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ В ТОРГОВЛЕ

Технический анализ и гипотеза эффективного рынка	193
Сбор данных и определение правил	196
Воспроизведение правила СМА нейронной сетью	201
Результаты работы нейронной сети	204
Обсуждение	208

СПИСОК ЛИТЕРАТУРЫ**ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ****От издательства**

	210
	230
	236

“Poverkh bar'ero v”¹

Так уж совпало: нам стало известно, что научное издательство «ТВП» проявило интерес к плоду наших раздумий — книге «Нейронные сети и финансовые рынки: принятие решений в торговых операциях» — как раз в те дни, когда компания Кристи оповестила книжных коллекционеров всего мира о том, что она выставляет на аукцион в Лондоне рукописи и подлинные письма Б. Пастернака.

Мы не понаслышке знаем, как высок в России уровень исследований в области адаптивных нелинейных моделей, и испытываем чувство гордости от того, что наша книга будет предложена вниманию столь квалифицированной читательской аудитории.

Почему эта книга была выбрана для перевода? Разумеется, мы можем лишь строить предположения на сей счет, но считаем нужным отметить, что в книге дается очень прагматичный и непредвзятый подход к трудным вопросам, возникшим в области корпоративных финансов после того, как на смену сегментации и регулированию пришли свободные финансовые потоки и свободный обмен достоверной информацией. Опыт работы в MBA-классах (Master of Business Administration) в Манчестерской Школе бизнеса (Великобритания) и Роттердамской Школе менеджмента (Голландия) убедил нас в том, что интеллектуалу-теоретику необходима скромность. Очень скоро нам стало ясно, что теоретическое превосходство той или иной модели ценообразования рынка ценных бумаг вовсе не гарантирует реального финансового успеха в виде роста счета в банке. Любая модель дает лишь общую схему для анализа сложных явлений реальной жизни и вместе с тем зачастую огорчает нас излишней жесткостью своих предположений. В то же время, финансовая информация распространяется не идеально, а на рынках происходят изменения, связанные с регулированием и конкуренцией.

Понимание этих обстоятельств заставило нас переключить свое внимание с универсальных моделей на новые перспективные методы. В 1990 году искусственные нейронные сети — в то время новая вещь для финансовой науки — преподносились как чудодейственное средство, способное почти полностью вытеснить человека из сферы операций на финансовых рынках. Буквально то же самое говорилось десятилетиями ранее про экспертные системы. В действительности, внутренняя негибкость, присущая системам, основанным на правилах и даже линейно согласованных соотношениях, не позволяет им быть адекватным средством анализа динамичных и из-

¹Б. Пастернак «Поверх барьеров», 1916.

менчивых рынков капитала. Поэтому когда нам в полной мере стали ясны потенциальные возможности нейронных сетей, мы почувствовали большое воодушевление. Искусственные нейронные сети, будучи методом, «свободным от моделей» ("model-free")², позволяют нам оценивать доход по финансовым активам или частоту невыполнения обязательств по контрактам, не связывая себя при этом ограничениями, которые налагаются в общепринятых теоретических моделях.

Более того, безмодельный подход позволил нам включать в рассмотрение в качестве потенциальных источников информации такие экзотические переменные, как погода (количество дождливых и солнечных дней), а также качественные переменные, скажем, организационную гибкость. Хотя, на первый взгляд, такие переменные кажутся слишком необычными, по крайней мере, интуитивно понятно, что они могут отвечать за какую-то составляющую дохода по активу или его волатильности. Так, в ходе работы в рамках нашего совместного проекта с Министерством финансов Голландии выявился ряд задач, где погодный фактор является высокосвязанной объясняющей переменной.³

Наши коллеги в финансовых подразделениях далеко не всегда приветствовали такой либеральный и неструктурный в смысле модели подход к проблеме. Примерно до 1994 года консервативные эконометристы отвергали саму идею о том, что нейронные сети могут быть средством моделирования. Сетевые методы подвергались насмешкам из-за того, что они устроены как «черный ящик», а также из-за наличия обучающих параметров, подбор которых производится практически «на глаз». В самом деле, разве профессор Х. Уайт не доказал со всей определенностью,⁴ что в задаче прогнозирования дохода по акциям IBM нейронные сети дают худшие результаты, чем модель случайного блуждания?

Позднее выяснилось, что данные по акциям IBM не являются достаточно репрезентативными, и поэтому данный результат ни в коем случае не компрометирует нейронно-сетевые методы. Постепенно сословие эконометристов искренне приняло нейронно-сетевую методологию как универсальное средство функциональной аппроксимации. В настоящее время фирмы-производители таких статистических программных пакетов, как SAS®, включают в них нейронно-сетевые модули.

Нам кажется, что сейчас интеллектуальная пыль несколько осела. Наша позиция предельно четкая: нейронные сети — не волшебная

палочка, а математический метод. Не будучи панацеей, нейронные сети позволяют, по крайней мере, отчасти, предсказать величину непредвиденного дохода по финансовым активам. В современном мире финансов, когда дневной оборот валютного рынка превышает 1 триллион долларов, улучшение качества прогноза всего на 1% принесет значительную прибыль. Далее, мы считаем, что наибольшую пользу нейронные сети могут принести в задачах обработки качественных данных. Европейская Комиссия включила в программу ESPRIT ряд исследовательских проектов, результаты которых позволили бы учитывать «новости» (мягкие данные) в моделях принятия финансовых решений. Прекрасным полем для дальнейшего развития этого направления может стать финансовый рынок России, уже сейчас играющий заметную роль в мире. Мы искренне надеемся на то, что наша книга будет способствовать лучшему пониманию возможностей нелинейных адаптивных методов, равно как и связанных с ними ловушек.

Мы очень благодарны проф. С.В. Курочкину (Вычислительный центр Российской Академии наук) за его кропотливый труд переводчика. Мы также глубоко признательны д-ру В.И. Хохлову (Математический ин-т им. В.А. Стеклова РАН) и издательству «ТВП» за энтузиазм и поддержку. К нашему великому сожалению, мы не знаем русского языка, и очень надеемся, что такие проекты, как этот, будут усиливать давление на западную систему образования с тем, чтобы ваш прекрасный язык был включен в учебные планы вузов.

Разумеется, все допущенные в книге ошибки остаются на нашей совести.

Д-р Дирк-Эмма Бюстенс

Женераль Банк

Брюссель

Д-р Виллем-Макс ван ден Берг

Факультет финансов

Университет Эразма, Роттердам

Проф. Д. Вуд

Манчестерская Школа Бизнеса

²Строго говоря, методов, «свободных от моделей», не существует!

³Более формальное доказательство этого факта см. в American Economic Review, 1994.

⁴Эта знаменитая работа перепечатана в [226] (см. список литературы).

Дирк-Эмма Бэстенс. С 1996 г. занимает должность управляющего по кредитному риску Женераль Банка, Брюссель, Бельгия. Отвечает за оптимизацию портфеля рискованных корпоративных кредитов. Предыдущая должность — ассистент-профессор финансового дела на экономическом факультете Университета Эразма в Роттердаме. Получил степень доктора философии (Ph.D.) в области экономики бизнеса в Манчестерской Школе бизнеса (Великобритания); тема диссертации: «Влияние регулирования рынка на финансовые институты». Изучал юриспруденцию в Католическом Университете г. Лувен и прикладную экономику в Университете Антверпена, где ему было присвоено звание магистра (Master of Science) в области морского права и навигации во внутренних водах.

Преподает курс «Корпоративные финансы и управление риском» в рамках программы MBA/MBI в Роттердамской Школе бизнеса. Публикуется в журналах: *Journal of International Securities Market*, *Neural Network World*, *European Journal of Finance*, *European Journal of Operational Research*.

Сведения об авторе книги приведены в изданиях «Кто есть кто в мире — 1996» и «Кто есть кто в мире финансов — 1997». Выступал в качестве консультанта по вычислительным методам в области финансов для Министерства финансов Голландии и банка ABN-AMRO.

Виллем-Макс ван ден Берг. Ассоциированный профессор по финансовому делу и инвестициям в Университете Эразма, Роттердам; совместно с банком ABN-AMRO разработал программный пакет для управляющих финансовыми активами, в котором используются теория игр и моделирование. Занимался исследованием возможностей нейронных сетей как инструмента для принятия решений в области финансов. В настоящее время является ответственным исполнителем проекта «ACE» («Рабочая среда аналитика») в рамках европейского проекта ESPRIT, где разрабатывает модели торговли в реальном времени, основывающиеся на информации, передаваемой по каналам новостей Рейтер.

Дуглас Вуд. Получил степень доктора философии (Ph.D.) в Университете Манчестера после обучения в Бирмингеме, Оксфорде и Массачусетском Технологическом институте. Поступил на работу на Экономический факультет Манчестерской Школы бизнеса, где в 1989 г. перешел на кафедру NatWest финансов и банковского дела. В течение многих лет был директором Международного Центра банковского дела и финансовых услуг при Манчестерской Школе бизнеса. Основная область его интересов — прогнозирование в финансах и в бизнесе. Он опубликовал 6 книг по этой теме, в том числе «Прогнозирование решений в бизнесе» и «Международный финансовый бизнес». Является редактором тома «Финансы» Новой Блэкуэлловской Энциклопедии менеджмента (1996), в настоящее время занимается вопросами влияния европейской денежной интеграции на характеристики риска и определение цены производных финансовых инструментов и на Европейскую платежную систему (последнему вопросу посвящена публикация «Европейская платежная система», *International Payments APACS* (ed.), London, 1996).

Повседневная практика финансовых рынков находится в интересном противоречии с академической точкой зрения, согласно которой изменения цен финансовых активов происходят мгновенно, без каких-либо усилий эффективно отражая всю доступную информацию. В действительности же, само существование сотен маркет-мейкеров, трейдеров и фондовых менеджеров, работа которых состоит в том, чтобы делать прибыль, говорит о том, что участники рынка вносят определенный вклад в общую информацию. Более того, так как эта работа стоит дорого, то и объем привнесенной информации должен быть значительным.

Труднее ответить на вопрос о том, как конкретно на финансовых рынках возникает и используется информация, которая может приносить прибыль. Исследования почти всегда показывают, что никакая устойчивая стратегия торговли не дает постоянной прибыли, и это, во всяком случае, так, если учитывать еще и расходы на совершение сделок. Хорошо известно также, что участники рынка (и весь рынок в целом) могут принимать совершенно различные решения, исходя из сходной или даже неизменной информации. Выход Великобритании из механизма валютных курсов европейской валютной системы (ERM) и октябрьский кризис 1987 г. — примеры ситуаций, когда трудно найти разумную объективную причину того, что данное событие произошло именно тогда, когда произошло, а не месяцем раньше или позже.

События такого рода свидетельствуют о том, что участники рынка в своей работе не ограничиваются линейными состоятельными правилами принятия решений, а имеют в запасе несколько сценариев действий, и то, какой из них пускается в ход, зависит подчас от внешне незаметных признаков.

Один из возможных подходов к многомерным и зачастую нелинейным информационным рядам финансового рынка заключается в том, чтобы по возможности подражать образцам поведения участников рынка, используя такие методы искусственного интеллекта, как экспертные системы или нейронные сети.

На моделирование процессов принятия решений этими методами было потрачено много усилий. Оказалось, однако, что экспертные системы в сложных ситуациях хорошо работают лишь тогда, когда системе присуща внутренняя стационарность (т.е. когда на каждый входной вектор имеется единственный не меняющийся со временем

ответ). Под такое описание в какой-то степени подходят задачи комплексной классификации или распределения кредитов, но оно представляется совершенно неубедительным для финансовых рынков с их непрерывными структурными изменениями. В случае с финансовыми рынками едва ли можно утверждать, что можно достичь полного или хотя бы в определенной степени адекватного знания о данной предметной области, в то время как для экспертных систем с алгоритмами, основанными на правилах, это — обычное требование.

Нейронные сети предлагают совершенно новые многообещающие возможности для банков и других финансовых институтов, которым по роду своей деятельности приходится решать задачи в условиях небольших априорных знаний о среде. Характер финансовых рынков драматическим образом меняется с тех пор, как вследствие ослабления контроля, приватизации и появления новых финансовых инструментов национальные рынки слились в общемировые, а в большинстве секторов рынка возросла свобода финансовых операций. Очевидно, что сами основы управления риском и доходом не могли не претерпеть изменений, коль скоро возможности диверсификации и стратегии защиты от риска изменились до неузнаваемости.

Стационарные линейные модели с большим трудом держатся на плаву в этом бурном море. Напротив, должным образом сконструированные нейронные сети, позволяющие определять по данным не только параметры, но и структуру системы, представляют собой весьма общую схему для описания развивающихся взаимосвязей.

Естественно, что различные области финансового дела вызывают к себе различный интерес. Так, например, одной из сфер применения нейронных сетей для ряда ведущих банков стала проблема изменений позиции доллара США на валютном рынке при большом числе неизменных объективных показателей. Возможности такого применения облегчаются тем, что имеются огромные базы экономических данных, — ведь сложные модели всегда прожорливы в отношении информации.

Котировки облигаций и арбитраж — еще одна область, где задачи расширения и сужения риска, разницы в процентных ставках и ликвидности, глубины и ликвидности рынка являются благоприятным материалом для мощных вычислительных методов.

Еще одной проблемой, значение которой в последнее время возрастает, является моделирование потоков средств между институциональными инвесторами. Падение процентных ставок сыграло решающую роль в повышении привлекательности инвестиционных

фондов открытого типа и индексных фондов, а наличие опционов и фьючерсов на их акции позволяет приобретать их с полной или частичной гарантией. Неудивительно, что волатильность (изменчивость) потоков наличности возросла и стратегии, основанные на правиле «занимая ненадолго, ссужай надолго», сделались более рискованными, а это, в свою очередь, привело к уменьшению денежных потоков через традиционные сберегательные институты. В итоге институты соревнуются за право вложить деньги, и все большее число менеджеров фондов и финансовых посредников имеет дело с международными рынками, в том числе и с такими их секторами, инвестирование в которые еще не так давно выглядело бы очень странно.

Очевидно, что задача оптимизации в условиях, когда число частичных ограничений равновесия бесконечно (например, на фьючерсном и наличном рынке любого товара в любом секторе рынка играют роль перекрестные разности процентных ставок), становится проблемой чрезвычайной сложности, все более выходящей за рамки возможностей любого трейдера.

В таких обстоятельствах трейдеры и, следовательно, любые системы, стремящиеся описать их поведение, в каждый момент времени должны будут сосредоточивать внимание на уменьшении размерности задачи. Хорошо известно такое явление, как ценная бумага повышенного спроса. Как показывают результаты анализа, проведенного Джеймсом Кейплом¹, весь прирост индекса FTA за последние 32 года был на самом деле достигнут за счет его повышений в течение только 25 месяцев из всего периода. Чем же так отличаются эти месяцы от остальных трехсот шестидесяти? Надо полагать, что линейная модель объяснить все это будет не в состоянии, да и в отношении нелинейной модели вопрос тоже остается открытым.

В любом случае приведенные далее результаты ряда независимых исследований подтверждают *априорное* предположение о том, что во временных рядах финансовых показателей присутствуют существенные нелинейности, и что нейронные сети — сами по себе или вместе с другими методами — могут сильно помочь в их распознавании.

Когда речь идет о финансовом секторе, можно с уверенностью утверждать, что первые результаты, полученные при применении нейронных сетей, являются весьма обнадеживающими, и исследования в этой области нужно развивать. Как это уже было с экспертными системами, может потребоваться несколько лет, прежде чем финансовые институты достаточно уверятся в возможностях нейронных сетей и станут использовать их на полную мощность. Финансовый мир очень сильно изменился за последнее десятилетие, и любые

¹Санди Таймс, 27 февраля 1994 г., с. 314

структурные подходы к моделированию неизбежно будут быстро устаревать, что делает бессмысленным их развитие, тогда как обучаемые системы типа нейронных сетей всегда будут приспосабливаться к переменам.

Др. Анна Папаиоанну
Отдел рынка долгосрочного ссудного капитала
Barclays de Zoete Wedd

Введение

В самое последнее время мы становимся свидетелями возникновения и первых испытаний на практике инвестиционных технологий нового поколения. Узкоспециализированные брокерские фирмы с экзотическими названиями вроде «Компания по предсказаниям» или «Пророческие системы» объединили свои усилия с такими финансовыми учреждениями, как Швейцарский банк, для проведения испытаний этих новых систем в условиях реального рынка.

Технологии, о которых идет речь, основываются на нелинейных методах анализа экономической и финансовой информации. В условиях возрастающей неуправляемости мировых процессов в финансовой сфере традиционные (читай, линейные) методы все чаще оказываются неспособными распознать ключевые переломы в тенденциях рынка. Так было, например, в случаях с крахом фондового рынка в 1987 году или началом глубокого спада в экономике Великобритании. Разочарование в этих методах заставило вспомнить о некогда казавшейся невероятной идее, согласно которой изменение рыночных показателей во времени не есть чисто случайное блуждание, а размеры ожидаемых доходов и/или характеристики неустойчивости (волатильности) можно пытаться находить при помощи более мощных методов. Общей чертой новых методов является возможность распознавания образов и вывода обобщающих правил. Существенными составными частями нового подхода являются: *нейронные сети* (сети компьютерных процессоров, взаимодействие которых построено по образцу процессов обучения, происходящих в человеческом мозге) и *генетические алгоритмы* (методы, в которых, исходя из большого набора первоначальных предположений, вырабатывают все более правильные представления о поведении рынка и, в конечном счете, более содержательные рабочие гипотезы). Про методы обоих видов говорят, что они управляются данными, в противоположность подходу, основанному на применении правил, который принят в *экспертных системах*. Системы, основанные на знаниях, обладают тем недостатком, что построенные на их основе методы торговли оказываются довольно негибкими. Наконец, совершенно новый взгляд на мир предлагает *теория динамических систем*, или теория хаоса. С ее помощью в явлениях, ранее считавшихся случайными, удастся обнаружить порядок или некоторую структуру. Основное предположение здесь состоит в том, что поведение системы есть результат множества нелинейных взаимодействий, вследствие чего даже небольшое изменение начальных данных может привести к совершенно другому дальнейшему поведению системы. Благодаря

достижениям этой теории, в некоторых ситуациях удастся обнаружить «дополнительные» доходы, которые нельзя было бы увидеть, действуя в предположениях гипотезы эффективного рынка (Efficient Market Hypothesis).

Эта книга посвящена нейронным сетям — мощному, хотя и не всегда правильно понимаемому методу. Говоря более конкретно, мы будем иметь дело с многослойными системами с алгоритмом обучения на основе обратного распространения ошибки (MBPN, Multilayer BackPropagation Network). Этот алгоритм наиболее широко распространен и наименее уязвим с точки зрения математики. Благодаря экспоненциально возрастающему в последние несколько лет количеству конференций по финансовым применениям нейронных сетей, эта методика завоевывает все большее доверие в финансовых кругах. Она имеет самые разные применения: от оптических устройств распознавания символов (OCR, Optical Character Recognition), основанных на нейронных моделях различения настоящих и поддельных подписей, до таких собственно финансовых приложений, как оценка кредитоспособности, процедуры рассмотрения заявок на ссуды, наилучшее распределение активов, оценка волатильности, управление валютными активами, технический анализ. Все эти темы будут предметом нашего рассмотрения в этой книге.

Данное направление исследований получило признание у практиков, так как оно согласуется с их интуитивными представлениями (а нередко и с выводами, которые делаются задним числом) о том, что в совокупной картине изменений показателей финансового рынка могут присутствовать определенные закономерности, которые можно распознать и на основе которых можно строить свою инвестиционную или торговую деятельность. Такие закономерности можно обнаруживать путем анализа временных рядов или анализа сечений. В любом случае применение нейронных моделей приносит ощутимую выгоду, и это прекрасно согласуется с тезисом Саймона об «ограниченной разумности», согласно которому на эффективности рынка сказывается ограниченность возможностей человека в работе с информацией. Более того, нейронные сети идеально приспособлены для обнаружения нелинейных зависимостей в отсутствие *априорных* знаний об основной модели. Нейронные сети можно использовать везде, где обычно применялись линейные методы (возможно, с предварительным преобразованием) и оценивание при помощи стандартных статистических методов.

Существуют и другие, более практические причины возрастания интереса к нелинейным методам анализа финансовых рынков. За последние несколько лет структура операций на рынках изменилась в сторону ее интернационализации; вместе с разнообразием в ставках коммиссионных пришли в беспорядок размеры сборов за совершение

сделок; сроки расчетов по совершенным сделкам сильно сократились; применение таких производных финансовых инструментов, как опционы, снизило ограничения по стоимости и капиталу при арбитражных сделках.

Когда мы рассматриваем сектор рынка, соответствующий долгосрочным сделкам, вероятно, следует признать, что определяющее влияние на цены оказывают здесь такие экзогенные факторы, как обменные курсы и процентные ставки, показатели экономического роста, тенденции (тренды) цен и показателей прибыли. Поведение рынка здесь удовлетворительно описывается так называемой гипотезой эффективного рынка, согласно которой в каждый момент вся доступная информация о текущих и будущих событиях дисконтируется в текущие цены рынка, так что изменения цен бывают вызваны только поступающей свежей информацией. Напротив, в краткосрочной перспективе появляются новые возможности для прогнозов, связанные с учетом регулирования платежей, обратных связей и многочисленных технических и структурных факторов. Старые парадигмы финансовой науки типа модели случайного блуждания или гипотезы эффективного рынка внушают нам представление о том, что финансовые рынки склонны относительно плавно и разумно приспосабливаться к поступающей информации. В этом круге идей вполне убедительно выглядят описания поведения рынка на основе линейных зависимостей и законов обращения трендов (стационарности). Однако драматические обвалы рынка при отсутствии существенных изменений информации, резкие изменения условий доступа и сроков при пересечении компаний какого-то невидимого порога в кредитной сфере — все это проявления нелинейности. Действительность показывает, что поведение финансовых рынков едва ли может быть описано линейными трендами.

Здесь, правда, необходимо некоторое предостережение. Преувеличенные претензии на сверхэффективность могут подорвать доверие к новейшим научным разработкам. На конференции «Нейронные сети и рынок долгосрочного ссудного капитала», организованной отделом нейропрогнозов Лондонской Школы бизнеса, вниманию участников были представлены нейронно-сетевые системы, якобы предсказывающие доходы по акциям с точностью 85%. При этом многое оставалось неясным, в частности: делается все это в рамках определенной модели или без каких-либо предположений; в расчете на доллары или в местной валюте; учитываются ли расходы на совершение сделок и т.д. Из-за недостатка информации такие результаты невозможно воспроизвести, и они не соответствуют уровню требований, сложившемуся к настоящему времени в эмпирической эконометрике. Быть может, все «восходящие звезды» финансовой науки — не более, чем кладоискатели. Вопреки заявлениям энтузиастов, имею-

щийся пока опыт носит двойственный характер. Даже узкие специалисты не сходятся во мнениях относительно того, куда приведут эти новые идеи — в тупик или, наоборот, к заветной чаше инвестиционного менеджмента — более высоким доходам без увеличения риска. Настоящую книгу не следует рассматривать как однозначную рекламу нейронных сетей. Чтобы правильно оценить результаты, полученные на основе нейронно-сетевого подхода, мы будем сравнивать их с тем, что дают в этом плане традиционные методы статистики — дискриминантный анализ, ARCH и регрессия.

Для проведения расчетов мы использовали несколько программных пакетов. Там, где требовались большие вычислительные мощности, применялся свободно распространяемый пакет Nevada Quickpropagation, разработанный Скоттом Фальманом в Университете Невады. Этот пакет без особых сложностей был откомпилирован для мэйнфрейм-компьютера Convex, установленного в Университете Эразма. Для менее громоздких задач использовался пакет ExploreNet 3000, разработанный Хехт-Нильсеном для среды Windows, а также транслятор C++ фирмы Symantech. Интересующиеся читатели могут обратиться к книге Блюма [44], которая представляет собой понятно написанное руководство по программированию нейронно-сетевых приложений на C++.

Книга написана таким образом, чтобы удовлетворить запросы как новичков, так и специалистов с опытом работы в области нейронных сетей. Мы всякий раз точно описываем финансовые принципы, лежащие в основе очередного приложения, при этом даются сведения из основ управления финансами, необходимые для полного понимания данного примера.

Первые две главы не содержат финансовых приложений и целиком посвящены основам нейронных сетей. В гл. 1 рассматриваются основные структуры и назначение нейронно-сетевых моделей. Описаны принципы разработки, обучения и оценки эффективности. Показано, каким образом множество задач, сильно различающихся параметрами сложности и устойчивости, может быть охвачено единой концепцией сети. В гл. 2 выясняется, насколько хорошо нейронные сети приспособлены для решения задач классификации и анализа временных рядов. Задача классификации понимается как задача отнесения предъявленного объекта к одному из нескольких попарно непересекающихся множеств. При этом наиболее важным случаем здесь является бинарная классификация — примерами ее могут служить распознавание доходных и не доходных инвестиций или различение компаний, имеющих хорошие шансы выжить, от тех, которые должны обанкротиться. В свою очередь, анализ временных рядов имеет целью определить будущие значения некоторой величины при

знании текущих и прошлых показателей ряда связанных с ней факторов.

В гл. 3 мы переходим к финансовым приложениям. Исследуется такой вопрос: обладают ли финансовые рынки внутренним механизмом нелинейной обратной связи? Если такой механизм, проявляющийся во внешне случайном, хаотическом поведении цен, действительно, существует, то это было бы серьезным ударом по таким известным теориям, как теория случайного блуждания или гипотеза эффективного рынка. При помощи ряда простых и хорошо известных статистических моделей мы исследуем некоторые временные ряды, а затем они же используются для определения возможностей нейронно-сетевого подхода в обнаружении (и предсказании) детерминированных связей в исследуемом ряде. В гл. 4 рассматриваются результаты управления активами и пассивами министерства финансов Голландии и, особенно подробно, ежемесячная оценка валовых поступлений от налогов. Оценивается эффективность различных методов, в том числе — модели ARIMA (собственной разработки министерства). Новые методы, такие, как нейронные сети, позволяют исследовать процесс без предварительной спецификации нелинейной модели, и, по-видимому, традиционные модели образования цен уступают им именно из-за отсутствия спецификации, а не из-за неучета свойств эффективности рынка.

В гл. 5 на примере Европейской биржи опционов (ЕОЕ, Амстердам) исследуется вопрос о том, дают ли нейронные сети существенные возможности для получения прибыли в течение одного торгового дня. Гл. 6 посвящена результатам определения макроэкономических показателей (так называемых глобальных (pervasive) факторов), влияющих на доходы от общего индекса акций на Нью-Йоркской и Амстердамской фондовых биржах. В гл. 7 показано, что нейронные сети являются вполне жизнеспособным инструментом отбора в международном распределении активов, поскольку они позволяют выбрать среди всех портфелей тот, который имеет наивысший доход и наименьший риск. Гл. 8 посвящена оценке кредитного риска посредством данных нефинансового характера. До сих пор очень мало было сделано для того, чтобы в моделях предсказания банкротств учитывать качественные показатели. В то же время, нейронные сети могут работать как с числовыми, так и с нечисловыми данными. В гл. 9, напротив, чисто экономические числовые показатели используются для оценки возможности банкротства корпорации на примере английских производителей комплектующих для автомобилей. В последней, десятой, главе построена нейронная сеть для обнаружения критических точек при изменении показателей доходов по акциям. Результаты показывают, что простое техническое правило торговли, реализованное нейронно-сетевой системой с прямой свя-

зью, способно сигнализировать о начинающемся росте доходности. В конце книги приведен довольно большой, но далеко не исчерпывающий, список литературы.

В заключение мы хотели бы поблагодарить тех, кто принимал активное участие в осуществлении этого проекта. В первую очередь, это — Эрве Водрей (факультет финансов Университета Эразма), внесший значительный вклад в содержание глав 1, 2 и 10. Его знания и опыт оказали неоценимую помощь в нашей работе. Бхаскар Дасгупта (Манчестерская Школа бизнеса) разработал описанную в гл. 7 нейронную сеть для предсказания национальных фондовых индексов. Кроме этого, финансовые факультеты Университета Эразма и Манчестерской Школы бизнеса предоставили нам все возможности для проведения исследований. Содержательные методы должны опираться на достоверные данные, и в этой связи мы хотели бы поблагодарить руководство Европейской биржи опционов в Амстердаме и Монику Дондерс за возможность доступа к базе биржевых данных по опционам. Мы благодарны Министерству финансов Голландии и, в особенности, Хану ван дер Кнопу и Франсу Хоймансу за данные о налоговых поступлениях и Хансу Рийнбергу из Нидерландского Инвестиционного банка (NIB) за сведения о качественных показателях оценки финансовой состоятельности инвесторов. Мы получили доступ к известной базе данных Чена, Ролла и Росса исключительно благодаря Вернеру де Бондту (Университет Висконсина).

Роттердам/Манчестер, январь 1994 г.

Нейронно-сетевые методы

В этой главе рассматриваются основные структуры и назначение нейронно-сетевых моделей. Описаны принципы разработки, обучения и оценки эффективности. Показано, каким образом множество задач, сильно различающихся параметрами сложности и устойчивости, может быть охвачено единой сетевой концепцией.

ВВЕДЕНИЕ В МЕТОДЫ НЕЙРОННЫХ СЕТЕЙ

Нейронные сети представляют собой новую и весьма перспективную вычислительную технологию, дающую новые подходы к исследованию динамических задач в финансовой области. Первоначально нейронные сети открыли новые возможности в области распознавания образов, затем к этому прибавились статистические и основанные на методах искусственного интеллекта средства поддержки принятия решений и решения задач в сфере финансов.

Способность к моделированию нелинейных процессов, работе с зашумленными данными и адаптивность дают возможность применять нейронные сети для решения широкого класса финансовых задач. В последние несколько лет на основе нейронных сетей было разработано много программных систем для применения в таких вопросах, как операции на товарном рынке, оценка вероятности банкротства банка, оценка кредитоспособности, контроль за инвестициями, размещение займов.

Приложения нейронных сетей охватывают самые разные области интересов: распознавание образов, обработка зашумленных данных, дополнение образов, ассоциативный поиск, классификация, составление расписаний, оптимизация, прогноз, диагностика, обработка сигналов, абстрагирование, управление процессами, сегментация данных, сжатие информации, сложные отображения, моделирование сложных процессов, машинное зрение, распознавание речи.

Смысл использования нейронных сетей в финансовой области заключается вовсе не в том, чтобы вытеснить традиционные методы или изобретать велосипед. Это просто еще одно возможное средство для решения задач, и цель этой книги — показать на целом ряде примеров, как можно применять нейронный подход и сравнивать его эффективность с эффективностью других методов. На основании этих примеров можно будет сделать общий вывод о том, что

нейронные сети довольно хорошо умеют отражать свойства разрывности, наблюдаемые в нашем мире. Неудивительно, что на исследования в этой области выделяются значительные средства: на 5-летнюю программу в США было выделено \$300,000,000, в Японии на 10-летнюю программу — \$400,000,000, финансирование в странах Европы составляет порядка \$100,000,000.

В этой главе, не углубляясь в детали теории, мы познакомим читателя-практика с основами сетевых вычислений. Более подробное изложение этих вопросов можно найти в многочисленной технической литературе. Наша цель сейчас — дать описание, что представляет собой нейронная сеть: ее составляющие, структура, конструктивные элементы, как понять логику ее работы, как описать ее возможное поведение. В следующей главе мы рассмотрим эти вопросы более конкретно для двух основных областей применения нейронных сетей — классификации и прогноза.

Параллели с биологией

Идея разработки систем обработки интеллектуальной информации по образу устройства нервной системы возникла давно. В 1943 г. МакКаллох и Питс создали упрощенную модель нервной клетки — *нейрон*. Мозг человека содержит до 10^{11} нейронов различных видов, при этом все они сложным образом связаны между собой и собраны в популяции — нейронные сети.

С биологической точки зрения, клетка состоит из *ядра*, *отростков* (дендритов), через которые информация поступает в клетку, и *аксона*, передающего выходной сигнал в другие клетки посредством тысяч разветвлений — *синапсов*. Простейший нейрон может иметь до 10^4 дендритов, принимающих электрохимические сигналы от других клеток. Определенные сложные комбинации этих входных сигналов, с учетом уровня чувствительности, вызывают возбуждение нейрона. После этого клетка через аксон передает сигнал другим клеткам, также имеющим свою систему связей. При поступлении сигнала изменяется вероятность возбуждения следующего нейрона. Если она увеличивается, то такая синаптическая связь называется *возбуждающей*, если уменьшается — *тормозящей*.

Внутри биологической клетки сигнал распространяется гораздо медленнее, чем в электронных схемах. Однако, вся сеть целиком оказывается высокоэффективной в решении таких сложных задач, как, например, распознавание образов (зрение, речь). Можно предположить, что причина таких потрясающих способностей мозга кроется в высокой организации связей и параллелизме в функционировании сети (см. [234]).

Характер разработок в области нейронных сетей принципиально отличается от экспертных систем: последние построены на утверждениях типа «если..., то...», которые нарабатываются в результате длительного процесса обучения системы, а прогресс достигается, главным образом, за счет более удачного использования формально-логических структур. В основе нейронных сетей лежит преимущественно поведенческий подход к решаемой задаче: сеть «учится на примерах» и подстраивает свои параметры при помощи так называемых алгоритмов обучения через механизм обратной связи.

УСТРОЙСТВО НЕЙРОННЫХ СЕТЕЙ

Различные виды искусственных нейронов

Искусственным нейроном называется простой элемент, сначала вычисляющий взвешенную сумму V входных величин x_i :

$$V = \sum_{i=1}^N W_i x_i = \vec{W} \cdot \vec{x}. \quad (1)$$

Здесь N — размерность пространства входных сигналов.

Затем полученная сумма сравнивается с пороговой величиной (или *bias*) W_0 , вслед за чем вступает в действие нелинейная функция активации f (ее можно также охарактеризовать как «решающую функцию»). Коэффициенты $\{W_i\}$ во взвешенной сумме (1) обычно называют *синаптическими коэффициентами* или *весами*. Саму же взвешенную сумму V мы будем называть *потенциалом* нейрона i . Выходной сигнал тогда имеет вид $f(V)$.

Величину порогового барьера можно рассматривать как еще один весовой коэффициент при постоянном входном сигнале. В этом случае мы говорим о *расширенном входном пространстве*: нейрон с N -мерным входом имеет $N+1$ весовой коэффициент. Если ввести в уравнение пороговую величину W_0 , оно переписывается так:

$$V = \sum_{i=1}^N W_i x_i + W_0 = \begin{pmatrix} \vec{W} \\ W_0 \end{pmatrix}^T \cdot \begin{pmatrix} \vec{x} \\ 1 \end{pmatrix}. \quad (2)$$

В зависимости от способа преобразования сигнала и характера функции активации возникают различные виды нейронных структур. Мы будем рассматривать только *детерминированные нейроны* (в противоположность *вероятностным нейронам*, состояние которых в момент t есть случайная функция потенциала и состояния в момент $t-1$). Далее, мы будем различать *статические нейроны* — такие, в которых сигнал передается без задержки, — и *динамические*, где

учитывается возможность таких задержек (учитывается («синапсы с запаздыванием»)).

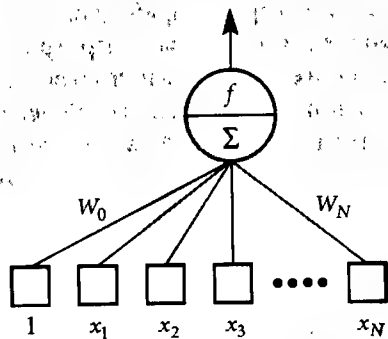


Рис. 1.1. Статический искусственный нейрон

Различные виды функции активации

Функции активации f могут быть различных видов:

- линейная: выходной сигнал нейрона равен его потенциалу,
- ступенчатая: нейрон принимает решение, выбирая один из двух вариантов (активен/неактивен),
- линейная с насыщением: нейрон выдает значения, промежуточные между двумя предельными значениями A и B ,
- многопороговая: выходной сигнал может принимать одно из q значений, определяемых $(q-1)$ порогом внутри предельных значений A и B ,
- сигмоидная: рассматриваются два вида сигмоидных функций:

$$s = f(V) = \frac{1}{1 + \exp(-bV)} \quad (3)$$

с выходными значениями в промежутке $(0,1)$ и

$$s = f(V) = \frac{\exp(bV) - 1}{\exp(bV) + 1} \quad (4)$$

с выходными значениями от -1 до 1 .

Коэффициент b определяет крутизну сигмоида. Поскольку сигмоидная функция является гладким отображением $(-\infty, +\infty) \rightarrow (0,1)$, крутизну b можно учесть через величины весов и порогов, и без ограничения общности можно полагать ее равной единице.

Возможно также определить нейроны без насыщения, принимающие на выходе непрерывное множество значений. В задачах классификации выходное значение может определяться порогом —

при принятии единственного решения, — или быть вероятностным — при определении принадлежности к классу. Чтобы учесть особенности конкретной задачи, могут быть выбраны различные другие виды функции активации — гауссова, синусоидальная, всплески (wavelets) и т.д.

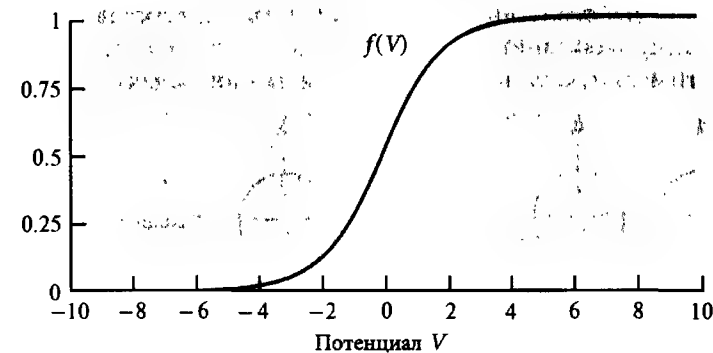


Рис. 1.2. Стандартная сигмоидная функция (крутизна = 1)

Нейронные сети с прямой связью

Искусственная нейронная сеть построена из нейронов, связанных друг с другом. Даже если работа нейронной сети имитируется на компьютере, лучше представлять ее себе не как программу, а как электронную схему. Мы будем рассматривать два вида нейронных сетей: статические, которые также часто называют сетями с прямой связью (feed-forward), и динамические, или рекуррентные сети. В этом разделе мы займемся статическими сетями. Сети других видов будут кратко рассмотрены позднее.

Нейронные сети с прямой связью состоят из статических нейронов, так что сигнал на выходе сети появляется в тот же момент, когда подаются сигналы на вход. Организация (топология) сети может быть различной. Если не все составляющие ее нейроны являются выходными, говорят, что сеть содержит скрытые нейроны. Наиболее общий тип архитектуры сети получается в случае, когда все нейроны связаны друг с другом (но без обратных связей). В конкретных задачах нейроны обычно бывают сгруппированы в слои. На рис. 1.3 показана типовая схема нейронной сети с прямой связью с одним скрытым слоем.

Интересно отметить, что, согласно теоретическим результатам [79], [118], нейронные сети с прямой связью и с сигмоидными функциями являются универсальным средством для приближения (аппроксимации) функций. Говоря точнее, любую вещественнозначную функцию нескольких переменных на компактной области определе-

ния можно сколь угодно точно приблизить с помощью трехслойной сети. При этом, правда, мы не знаем ни размеров сети, которая для этого потребуется, ни значений весов. Более того, из доказательства этих результатов видно, что число скрытых элементов неограниченно возрастает при увеличении точности приближения. Сети с прямой связью, действительно, могут служить универсальным средством для аппроксимации, но нет никакого правила, позволяющего найти оптимальную топологию сети для данной задачи.

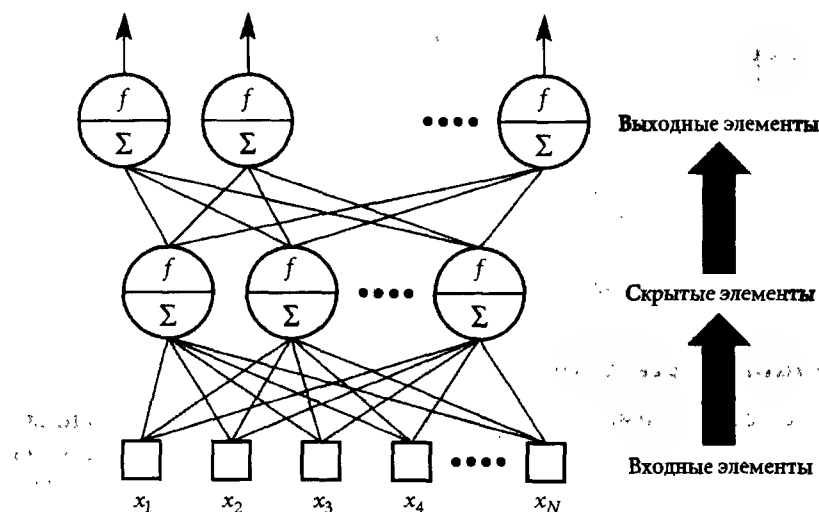


Рис. 1.3. Нейронная сеть с прямой связью с одним скрытым слоем (пороги не отмечены)

Таким образом, задача построения нейронной сети является нетривиальной. Вопросы о том, сколько нужно взять скрытых слоев, сколько элементов в каждом из них, сколько связей и какие обучающие параметры, в имеющейся литературе, как правило, трактуются облегченно. Предлагаемая обычно процедура состоит в переборе различных архитектур с целью найти «наилучшую» или хотя бы удовлетворительную модель. Такая концепция представляется весьма ограниченной.

В литературе можно найти многочисленные «правила большого пальца» для определения числа скрытых узлов или количества весовых коэффициентов. В качестве рецептов могут предлагаться: среднее арифметическое числа входов и выходов, половина числа входов или удвоенный квадратный корень из произведения входов и выходов. Такие предложения математически необоснованны, и исследования в этой области ведутся очень активно.

Каждый узел многослойной сети проектирует свой входной вектор на вектор весов посредством скалярного произведения. Таким

образом, в простейшем случае сети с одним скрытым слоем вектор входного сигнала проектируется на вектор весов каждого из нейронов скрытого слоя. В результате мы получаем иное представление входного сигнала. После этого результаты проекций подвергаются нелинейным преобразованиям. Их цель — усилить те характеристики, за которые отвечает соответствующий узел. Компоненты вектора весов каждого узла подобраны так, чтобы выделить определенные свойства сигнала. Посредством нелинейного преобразования способность узла улавливать эти свойства увеличивается. Число характеристик, которые нужно уловить, чтобы осуществить правильную конструкцию сети, зависит от исходной задачи.

Выбор вариантов схемы сети может быть сужен также за счет *априорных* знаний о задаче. Так, Ле Куном [175] была разработана сеть с «распределенными весами» для цифрового распознавания рукописного текста (где разные области входного изображения накрывались одинаковыми масками, извлекающими нужные характеристики).

Итак, мы описали, каким образом сеть с прямой связью осуществляет преобразование сигнала посредством весов, приписанных соединениям. Но как правильно выбрать значения этих весов? Этот момент обычно называют «обучением» или «тренировкой сети». Мы займемся им в следующем разделе.

ОБУЧЕНИЕ

На этапе обучения происходит вычисление синаптических коэффициентов в процессе решения нейронной сетью задач (классификации, предсказания временных рядов и др.), в которых нужный ответ определяется не по правилам, а с помощью примеров, сгруппированных в обучающие множества. Такое множество состоит из ряда примеров с указанным для каждого из них значением выходного параметра, которое было бы желательно получить. Действия, которые при этом происходят, можно назвать *контролируемым обучением*: «учитель» подает на вход сети вектор исходных данных, а на выходной узел сообщает желаемое значение результата вычислений. Контролируемое обучение нейронной сети можно рассматривать как решение оптимизационной задачи. Ее целью является минимизация функции ошибок, или невязки, E на данном множестве примеров путем выбора значений весов W . Суть задачи оптимизации станет яснее, если представить себе график невязки, рассматриваемой как функция весов (эта функция определена в многомерном пространстве весов, где каждому весовому коэффициенту соответствует своя размерность). Из-за нелинейностей функций активации полученная поверхность в общем случае будет очень сложной: наряду с плоскими участками на ней будут локальные минимумы, седловые точки,

овраги. На рис. 1.4 видно, что даже простая сеть может реализовывать сложный нелинейный процесс.

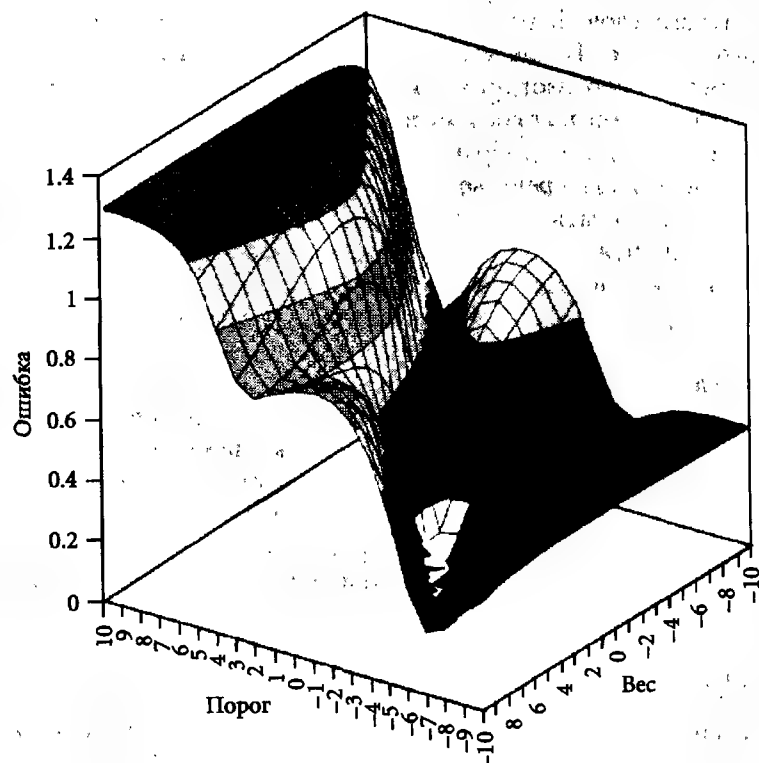


Рис. 1.4. Поверхность невязки для сети с одним узлом (1 вес и 1 порог) и сигмоидной функцией активации. Были взяты 3 примера со входами 0.6, 0.7, -0.3 и значениями выхода, соответственно, 0.5, 0.25 и 0.3

Критерии ошибок

Целью процедуры минимизации является отыскание глобального минимума — достижение его называется *сходимостью* процесса обучения. Поскольку невязка зависит от весов нелинейно, получить решение в аналитической форме невозможно, и поиск глобального минимума осуществляется посредством итерационного процесса — так называемого *обучающего алгоритма*, который исследует поверхность невязки и стремится обнаружить на ней точку глобального минимума. Иногда такой алгоритм сравнивают с кенгуром, который хочет попасть на вершину Эвереста, прыгая случайным образом в разные стороны. Разработано уже более сотни разных обучающих алгоритмов, отличающихся друг от друга стратегией оптимизации и критерием ошибок.

Коль скоро обучение основывается на минимизации значения некоторой функции (показывающей, насколько результат, который выдает сеть на данном обучающем множестве, далек от образцового значения), нужно, прежде всего, выбрать меру ошибки, соответствующую сути задачи. Удачный выбор меры погрешности обычно приводит к более гладкой поверхности невязки и упрощает задачу обучения (см. [269]). Обычно в качестве меры погрешности берется средняя квадратичная ошибка (MSE), которая определяется как сумма квадратов разностей между желаемой величиной выхода d_k и реально полученными на сети значениями y_k для каждого примера k :

$$E = \frac{1}{P} \sum_{k=1}^P (d_k - y_k)^2. \quad (5)$$

Здесь P — число примеров в обучающем множестве.

Наряду с такой мерой погрешности широко используется расстояние Кульбака–Лейблера, связанное с *критерием максимума правдоподобия*:

$$E = \sum_k \left[d_k \log \frac{d_k}{y_k} + (1 - d_k) \log \frac{1 - d_k}{1 - y_k} \right], \quad (6)$$

а также некоторые другие.

Минимизация величины E осуществляется с помощью градиентных методов. В первом из них берется градиент общей ошибки, и веса W пересчитываются каждый раз после обработки всей совокупности обучающих примеров («эпохи»). Изменение весов происходит в направлении, обратном к направлению наибольшей крутизны для функции стоимости:

$$W(t+1) = W(t) - \varepsilon \frac{\partial E}{\partial W}. \quad (7)$$

Здесь ε — определяемый пользователем параметр, который называется *величиной градиентного шага* или *коэффициентом обучения*.

Другой возможный метод носит название *стохастического градиентного*. В нем веса пересчитываются после каждого просчета всех примеров из одного обучающего множества, и при этом используется частичная функция стоимости, соответствующая этому, например, k -му, множеству:

$$W(t+1) = W(t) - \varepsilon \frac{\partial E_k}{\partial W}. \quad (8)$$

Обратное распространение ошибки

Рассмотрим теперь наиболее распространенный алгоритм обучения нейронных сетей с прямой связью — *алгоритм обратного распространения ошибки* (Backpropagation, BP), представляющий собой развитие так называемого *обобщенного дельта-правила* (см. [281]). Этот алгоритм был заново открыт и популяризирован в 1986 г. Румельхартом и МакКлелландом из знаменитой Группы по изучению параллельных распределенных процессов в Массачусетском технологическом институте. В этом пункте мы более подробно рассмотрим математическую суть алгоритма. Он является алгоритмом градиентного спуска, минимизирующим суммарную квадратичную ошибку:

$$E = \frac{1}{2} \sum_{k=1}^P \sum_i (d_k^i - y_k^i)^2. \quad (9)$$

Здесь индекс i пробегает все выходы многослойной сети.

Основная идея BP состоит в том, чтобы вычислять чувствительность ошибки сети к изменениям весов. Для этого нужно вычислить частные производные от ошибки по весам. Пусть обучающее множество состоит из P образцов, и входы k -го образца обозначены через $\{x_k^i\}$. Вычисление частных производных осуществляется по *правилу цепи*: вес входа i -го нейрона, идущего от j -го нейрона, пересчитывается по формуле

$$\Delta W_{ij} = -\epsilon \sum_{k=1}^P \frac{\partial E_k}{\partial W_{ij}} = -\epsilon \sum_{k=1}^P \frac{\partial E_k}{\partial v_k^i} \frac{\partial v_k^i}{\partial W_{ij}} = -\epsilon \sum_{k=1}^P \delta_k^i x_k^i, \quad (10)$$

где ϵ — длина шага в направлении, обратном к градиенту.

Если рассмотреть отдельно k -й образец, то соответствующее изменение весов равно

$$\Delta W_{ij} = -\epsilon \frac{\partial E_k}{\partial W_{ij}} = -\epsilon \delta_k^i x_k^i. \quad (11)$$

Множитель δ_k^i вычисляется через аналогичные множители из последующего слоя, и ошибка, таким образом, передается в обратном направлении.

Для выходных элементов мы получаем:

$$\delta_k^i = -\frac{\partial E_k}{\partial v_k^i} = -\frac{\partial E_k}{\partial x_k^i} \frac{\partial x_k^i}{\partial v_k^i} = -(d_k^i - x_k^i) f'(v_k^i). \quad (12)$$

Для скрытых элементов множитель δ_k^i определяется так:

$$\delta_k^i = -\frac{\partial E_k}{\partial v_k^i} = -\sum_h \frac{\partial E_k}{\partial v_k^h} \frac{\partial v_k^h}{\partial v_k^i}. \quad (13)$$

С учетом того, что

$$\frac{\partial E_k}{\partial v_k^h} = \delta_k^h \quad \text{и} \quad \frac{\partial v_k^h}{\partial v_k^i} = W_{hi} f'(v_k^i), \quad (14)$$

получаем:

$$\delta_k^i = -\frac{\partial E_k}{\partial v_k^i} = -f'(v_k^i) \sum_h W_{hi} \delta_k^h, \quad (15)$$

где индекс h пробегает номера всех нейронов, на которые воздействует i -й нейрон.

Данный алгоритм используется в двух вариантах. В *стохастическом* варианте веса пересчитываются каждый раз после просчета очередного образца, а в «эпохальном», или off-line варианте, веса меняются после просчета всего обучающего множества.

Способы обеспечения и ускорения сходимости

Выбор начальных весов

Перед тем, как начинать процесс обучения нейронной сети, необходимо присвоить весам начальные значения. Цель здесь, очевидно, должна состоять в том, чтобы найти как можно более хорошее начальное приближение к решению и таким образом сэкономить время обучения и улучшить сходимость. Конечно, можно положить начальные веса во всей сети равными нулю, но тогда частные производные от невязки по всем весам будут одинаковыми, и изменения весов не будут должным образом структурированы. В результате нельзя будет надеяться на то, что сеть вообще когда-нибудь сможет решить задачу. Нужно искать способы уйти от такой симметрии.

Классический подход к проблеме выбора начальных значений весов состоит в следующем: случайным образом выбрать малые величины весов, чтобы быть уверенным, что ни один из сигмоидных элементов не насыщен (и значения всех производных очень малы). Например, если сигмоидный элемент имеет наибольшую производную в интервале $[-3, 3]$, подходящим выбором для начального веса будет случайное число, расположенное в интервале $[-3/\sqrt{N}, 3/\sqrt{N}]$, где N — число входов нейрона. Из-за малости величины начального веса последующие его изменения также будут небольшими, так как при этом учитывается абсолютная величина веса. Таким образом, нужно найти компромисс между существенными изменениями в значениях сигмоидов и малостью абсолютных величин.

Подобные правила рассчитаны на то, чтобы сеть начинала свою работу в линейном режиме и притом не на плоской части поверхности невязок. Однако нет гарантии, что такое начальное приближение приведет к глобальному минимуму или уменьшит время сходимости. Были разработаны другие методы, дающие еще более хорошее начальное приближение с точки зрения уменьшения времени обучения и обладающие большей устойчивостью в смысле локальных минимумов. Так, Дено и Ланжель разработали метод инициализации весов по прототипам, полученным из обучающего множества [87]. Усовершенствованный классический метод выбора начальных значений использует данные анализа главных компонент, но для этого, безусловно, требуется меньше скрытых элементов, чем имеется входов [292]. При использовании обучающих алгоритмов типа ВР выбор начального приближения очень важен. Уже на этом шаге нужно позаботиться о том, чтобы не попасть в локальный минимум.

Обход локальных минимумов

Как уже говорилось, поверхность невязки в пространстве весов в общем случае имеет локальные минимумы, и это является главным препятствием для процесса обучения нейронной сети, в особенности, для алгоритма спуска. Можно встретить утверждения, что в ряде случаев локальный минимум является вполне приемлемым решением [105], однако в общей ситуации необходимо разработать стратегию, которая позволяла бы избегать таких точек и гарантировала бы сходимость обучающего алгоритма к глобальному решению.

Упорядочение данных

Для того чтобы обучающий алгоритм не стал двигаться в ложном направлении, нужно, прежде всего, упорядочить случайным образом последовательность примеров, которые он обрабатывает (так называемое «взбалтывание»). Это не так существенно для «эпохального» варианта ВР, но очень важно при стохастическом определении весов. Всякая программная реализация алгоритма типа ВР должна включать в себя случайное упорядочение обучающего множества. Более того, если какой-то из классов примеров представлен недостаточно, случайный выбор должен осуществляться таким образом, чтобы примеры из слабо представленной группы встречались чаще — этим будет устранен ложный крен при минимизации невязки. В гл. 8 читатель найдет иллюстрирующий пример.

Пакетная обработка

При использовании *стохастического варианта* ВР (когда веса изменяются после каждого примера) может получиться так, что сеть будет зря тратить время на перемещения туда-обратно. Даже в про-

стой задаче двоичной классификации на одном нейроне гиперплоскость может много раз колебаться вокруг своего оптимального положения, пока алгоритм будет обрабатывать поочередно образцы из разных классов. От этого явления несложно избавиться, взяв суммарное изменение весов после ряда примеров. Такой прием называется *пакетной* обработкой. Цель ее состоит в том, чтобы вычислять усредненное изменение весов. Очевидно, «*эпохальный*» вариант ВР представляет собой ту же пакетную обработку, где пакетом является все обучающее множество. Если в начале действия алгоритма брать небольшие пакеты, а затем увеличивать их объем вплоть до общего количества примеров в обучающем множестве, то этим будет сэкономлено время вычислений, но, тем не менее, обеспечена сходимость к глобальному решению. Особенно заметным этот эффект будет, если общее количество примеров очень велико или их размерность большая.

Импульс

Другой часто применяемый прием состоит в том, что при определении направления поиска к текущему градиенту добавляется поправка — вектор смещения с предыдущего шага, взятый с некоторым коэффициентом. Можно сказать, что учитывается уже имеющийся импульс движения. Окончательная формула для изменения весов выглядит так:

$$\Delta W_{ij}(t+1) = \mu \Delta W_{ij}(t) - (1-\mu) \varepsilon \frac{\partial E}{\partial W_{ij}}, \quad (16)$$

где μ — число в интервале (0,1), которое задается пользователем.

Часто значение μ волевым образом задается равным 0.9, безотносительно к специфике задачи и архитектуре сети. Нужно отметить, что метод импульса очень чувствителен к способу упорядочения примеров в обучающем множестве. Если подряд попало несколько примеров из одного класса, импульс станет очень большим, и это нарушит ход процесса обучения. Сказанное еще раз свидетельствует о необходимости предварительного случайного упорядочения обучающего множества.

Управление величиной шага

Ранее уже было сказано, что выбор величины шага имеет ключевое значение для успешной работы обучающего алгоритма. При слишком маленьком шаге обучение будет медленным и велика будет вероятность попадания в локальный минимум на поверхности невязки. Наоборот, при слишком большом шаге можно проскочить мимо глобального минимума. То, какой шаг следует считать малым,

а какой большим, сильно зависит от архитектуры сети и от обучающих данных. Есть различные эвристические соображения для выбора шага. Можно уменьшать шаг по мере хода итераций, или же нормировать его на евклидову норму градиента невязки с предыдущего шага. Можно поступать и так: если на предыдущей итерации ошибка уменьшилась, величина шага умножается на некоторый коэффициент, больший единицы («поощрение»), а если увеличилась, то на коэффициент, меньший единицы («наказание»). Нужно понимать, что такого рода правила выбора шага являются чисто эвристическими и не могут гарантировать улучшение сходимости.

Изменение производной сигмоида

Другой давно известный вычислительный трюк состоит в том, чтобы изменить величину производной сигмоида, умножив его на некоторую малую величину (обычно 0.1) с целью предотвратить «прилипание» к плоскому участку. Опыт, однако, показывает, что этот прием следует применять с осторожностью, так как он может породить «вынужденную» сходимость в неправильном направлении.

Методы второго порядка

При более последовательном подходе для улучшения процесса обучения можно использовать информацию о производных второго порядка от функции невязки. Соответствующие методы оптимизации называются квадратичными. Вся указанная информация собрана в матрице гессиана H , имеющей размеры $Nw \times Nw$, где Nw — число весов. Эта матрица содержит информацию о том, как изменяется градиент при малых смещениях по различным направлениям в пространстве весов. Прямое вычисление матрицы требует большого времени, поэтому разработаны методы, позволяющие избежать вычисления и хранения матрицы (спуск по сопряженному градиенту, масштабированный метод сопряженных градиентов (см. [197]), RBackProp (см. [212]), квази-ньютоновский метод, метод Левенберга-Маркара).

Методы локальной оптимизации

В отличие от методов второго порядка, где веса изменяются пропорционально их вкладу в направление глобального поиска, в локальных методах оптимизации каждый вес меняется локально. В качестве примера таких методов можно назвать метод дельта-дельта (см. [151]), Rprop (см. [240]), QuickProp (см. [104]).

Другие алгоритмы обучения

Многочисленные исследования посвящены сравнению различных обучающих эвристик (см., например, [104]). В большинстве имеющихся нейронно-сетевых пакетов реализованы методы пакетной обработки, импульса, изменения величины шага, и даже более совершенные варианты алгоритмов типа алгоритма обратного распространения ошибки и квази-ньютоновские методы. В литературе описано много других алгоритмов, реализующих иные подходы к задаче оптимизации. Так, в основанном на идеях статистической физики методе «замораживания» стабилизация алгоритма осуществляется за счет понижения «температурного» параметра. Другие подходы, такие, как метод случайного блуждания [13] или Alorex [269] используют случайный поиск в пространстве весов, и это принципиально отличает их от систематического поиска в методе обратного распространения ошибки (см. [78]). Наконец, в последнее время пользуются успехом так называемые *генетические алгоритмы*, в которых набор весов рассматривается как индивид, подверженный *мутациям* и *скрещиванию*, а в качестве показателя его «качества» берется критерий ошибки. По мере того, как нарождаются новые поколения, все более вероятным становится появление оптимального индивида.

Выбор эффективного обучающего алгоритма всегда включает в себя компромисс между сложностью решаемой задачи и техническими ограничениями (быстродействие и объем памяти компьютера, время, цена).

ОБОБЩАЮЩИЕ ПРАВИЛА

В предыдущем разделе мы описали различные способы приблизить функцию, которую реализует реальная сеть, к неизвестной функции, которую, как предполагается, можно определить по имеющемуся множеству примеров — обучающему множеству. Как в задачах классификации, так и в задачах прогноза, цель при построении сети должна состоять не в том, чтобы *запомнить* обучающую информацию, а в том, чтобы на основании изучения прошлого сделать определенные *обобщения*, которые можно будет затем применить к новым образцам. В конечном счете, эффективность сети определяется тем, как она работает со всей совокупностью возможных примеров (пространством возможных ситуаций). Так как все это множество целиком, как правило, недоступно, возникает практическая задача максимизации качества работы сети на всем множестве исходных данных, и для этого вовсе не нужно требовать от сети высокой степени соответствия на каком-то «зашумленном» обучающем множестве.

Шум

В каждой реальной задаче присутствует шум, и необходимо уметь справляться с ним. В особенности это относится к задачам обработки временных рядов, в которых переменные получены в результате измерений в некоторой физической системе, причем в самой системе и/или в механизме измерений шум присутствует естественным образом. В финансовых приложениях данные зашумлены особенно сильно. Например, совершение сделок может регистрироваться в базе данных с запозданием, причем в разных случаях — с разным. Пропуск значений или неполную информацию также иногда рассматривают как шум: в таких случаях берется среднее или наилучшее значение, и это, конечно, приводит к зашумлению базы данных. Отрицательно сказывается на обучении неправильное определение класса объекта в задачах распознавания — это ухудшает способность системы к обобщению при работе с новыми (т.е. не входившими в число образцов) объектами.

Переобучение

Как отмечалось выше, нейронные сети могут служить универсальным средством аппроксимации в том смысле, что при достаточно разветвленной архитектуре они реализуют широкий класс функций [79]. Как часто бывает, достоинство одновременно является и недостатком. Благодаря способности тонко улавливать структуру аппроксимируемой функции сеть достигает очень высокой степени соответствия на обучающем множестве, и в результате плохо делает обобщения при последующей работе с реальными данными. Это явление называется *переобучением*, или *эффектом бабушкиного воспитания*. Сеть моделирует не столько саму функцию, сколько присутствующий в обучающем множестве шум. Переобучение присутствует и в таких более простых моделях, как линейная регрессия, но там оно не так выражено, поскольку через обучающие данные нужно провести всего лишь прямую линию. Чем богаче набор моделирующих функций, тем больше риск переобучения. На рис. 1.5 показаны типичные проявления переобучения.

В следующем разделе мы рассмотрим некоторые приемы, позволяющие сети избежать чересчур точного следования обучающим примерам. В частности, нужно уметь распознать момент, когда обучение становится излишне точным. Вопросы, касающиеся переобучения и слабой способности к обобщению, связаны с общей проблемой отделения сигнала от шума.

Объем обучающей выборки

Первое естественное желание состоит в том, чтобы увеличивать число примеров в обучающем множестве. Чем их больше, тем более *представительны* данные. Как и в любом физическом измерении, увеличение числа наблюдений уменьшает шум. Если имеется несколько измерений одного объекта, сеть возьмет их среднее значение, и это лучше, чем точно следовать одному единственному зашумленному значению.

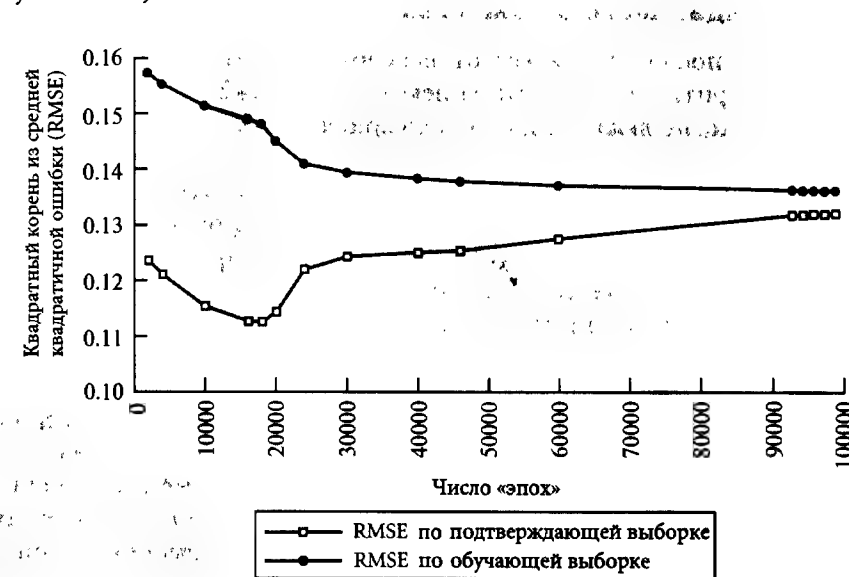


Рис. 1.5. Типичный случай переобучения

Однако на практике и, особенно, в финансовых приложениях невозможно получить такое количество наблюдений, которое было бы желательно в свете положений статистики. Число необходимых примеров резко растет с увеличением сложности моделируемой функции и повышением уровня шума. Более того, доступные нам данные могут иметь все меньшее отношение к делу. Как, например, информация, собранная в 1950 году, может быть значимой для описания современного положения в финансовом мире? Наконец, могут существовать физические ограничения на размер базы данных, например, объем памяти или недопустимо большое время обучения.

Вопросам, связанным с объемом множества образцов и сложностью сети, посвящены многочисленные исследования. В частности, изучались пороговые требования, которым должна удовлетворять система классификации, чтобы быть линейно отделяющей (см. [76]). Более точные результаты получены для пороговых нейронов: Баум

установил, что на сети с прямой связью, построенной из линейных пороговых функциональных элементов, можно получить правильные обобщения, если объем обучающего множества в несколько раз больше объема сети. Для многослойных сетей общего вида, построенных из сигмоидальных элементов, аналогичное утверждение не имеет места (см. [31]).

Последовательный спуск, или использование подтверждающего множества

Другой способ избавиться от переобучения заключается в том, чтобы измерить ошибку сети на некотором множестве примеров из базы данных, не включенных в обучающее множество, — *подтверждающем* множестве. Ухудшение характеристик сети при работе с этим множеством указывает на возможное переобучение. Наоборот, если характеристики улучшаются, это значит, что обучение продолжается. Таким образом, переобучение можно обнаружить, наблюдая за тем, насколько *последовательно уменьшается* ошибка во время обучения сети. В любом реальном (не смоделированном на компьютере) приложении нужно использовать подтверждающее множество, так как уровень шума заранее не известен.

Недостатком этого приема является уменьшение числа примеров, которые можно было бы взять в качестве обучающего множества. Малость базы данных — это серьезная проблема. Более того, оценка качества работы сети зависит от выбора образцов, составляющих подтверждающее множество. Даже при случайной выборке разные разбиения базы данных на обучающее и подтверждающее множества дают разные оценки. При исследовании этой проблемы Де Гроот [84] использовал для отбора примеров и построения обучающего и подтверждающего множеств алгоритмы кластеризации.

Перекрестное подтверждение

Для того чтобы устранить произвол в разбиении базы данных, могут быть применены методы повторных проб (см. [262]). Рассмотрим один из таких методов, который называется *перекрестным подтверждением*. Его идея состоит в том, чтобы случайным образом разбить базу данных на q попарно не пересекающихся подмножеств. Затем производится q обучений на $(q-1)$ множестве, а ошибка вычисляется по оставшемуся множеству. Если q достаточно велико, например, равно 10, каждое обучение задействует большую часть исходных данных. Если процедура обучения надежна, то результаты по q различным моделям должны быть очень близки друг к другу. После этого итоговая характеристика определяется как среднее всех по-

лученных значений ошибки. К сожалению, при применении этого метода объем вычислений часто оказывается очень большим, так как требуется проделать q обучений, и в реальном приложении с большой размерностью это может быть невыполнимо. В предельном случае, когда $q = P$, где P — общее число примеров, метод называется *перекрестным подтверждением с одним в остатке*. Такой метод оценки имеет смещение, и разработан метод «*складного ножа*», уменьшающий этот недостаток ценой еще большего объема вычислений.

Регуляризация

Еще один способ избежать переобучения состоит в том, чтобы ограничить совокупность функций отображения, реализуемых сетью. Методы такого типа называются *регуляризацией*. Например, в функцию стоимости может быть добавлено штрафное слагаемое, подавляющее резкие скачки отображающей функции (на математическом языке — большие значения ее второй производной). Алгоритм обучения изменяется таким образом, чтобы учитывался этот штраф (см. [126]).

Оптимизация архитектуры

Разработаны различные методы изменения архитектуры сети с целью повысить способность сети к обобщению. Здесь есть два основных подхода:

- **деструктивный подход:** берется сеть заведомо большего размера, чем нужно, и в процессе обучения из нее удаляются связи и даже сами нейроны;
- **конструктивный подход:** первоначально берется маленькая сеть, и к ней, в соответствии со структурой и сложностью задачи, добавляются новые элементы.

Примером деструктивного подхода является метод *уменьшения весов* (он похож на то, что статистики называют *гребневой регрессией*). Перлмутер был первым, кто применил его для того, чтобы предотвратить чрезмерный рост весов. Он включил в функцию стоимости штрафное слагаемое:

$$E = \sum_{k=1}^P E_k + \frac{\eta}{2} \sum_{i \rightarrow j} W_{ij}^2 \quad (17)$$

с $\eta < 1$. Заметьте, что множитель η можно считать отношением среднего квадратичного остатков к среднему квадратичному весов. Добавленный член вызовет такое изменение весов:

$$W_{ij} = (1 - \eta)W_{ij}, \quad (18)$$

заставляя уменьшаться те веса, на которые не действует первый член. Очевидно, чем больше вес, тем большее влияние он оказывает на функцию стоимости. Во втором варианте выражение для штрафа берется в виде

$$E = \sum_{k=1}^p E_k + \frac{\eta}{2} \sum_{i \rightarrow j} \frac{W_{ij}^2}{1 + W_{ij}^2}. \quad (19)$$

В результате малые коэффициенты убывают быстрее, чем большие. Кроме того, уменьшение весов помогает уходить с плоских участков поверхности на ранних стадиях обучения. Были предложены и другие виды выражений для штрафа, в результате чего удаляются не только соединения, но и нейроны (см. [132], [64]). Еще один метод уменьшения числа связей — «минимизация вреда для мозга» (см. [174]). Цель его состояла в том, чтобы находить в сети те веса, которые можно удалить, не меняя существенно среднеквадратичную ошибку (MSE) на обучающем множестве. Вводится показатель s_i (так называемая «выпуклость» веса) по формуле:

$$s_i = \frac{1}{2} \frac{\partial^2 \text{MSE}}{\partial W_i^2} W_i. \quad (20)$$

Удаление весов с малыми выпуклостями и повторное обучение урезанной сети улучшают ее общие характеристики. При итеративном применении этого метода к многослойному персептрону в задаче распознавания рукописного текста из сети было удалено более 50 процентов связей, и это привело к существенному уменьшению доли неправильно опознанных объектов (см. [174]).

В некоторых конструктивных методах наращивание сети происходит одновременно с обучением, см., например, [17], STEPNET [164], «черепичный» алгоритм [194], «всплеск» [116], нейронные деревья [244], каскадная корреляция [106].

Динамические, самоорганизующиеся сети и сети со встречным распространением

Нейронные сети с прямой связью и обучением методом обратного распространения ошибки рассматриваются в литературе чаще других. Кроме них, существует много других сетевых моделей, некоторые из которых имеют вычурные названия: «конкурентное обучение» (или «адаптивная теория резонанса»), сети Хопфилда, машины Больцмана, самоорганизующиеся карты признаков Кохонена.

Динамические сети

Следующий класс нейронных сетей, который мы рассмотрим, — динамические, или рекуррентные, сети. Они построены из динамических нейронов, чье поведение описывается дифференциальными или разностными уравнениями, как правило, — первого порядка. Сеть организована так, что каждый нейрон получает входную информацию от других нейронов (возможно, и от себя самого) и из окружающей среды. Этот тип сетей имеет важное значение, так как с их помощью можно моделировать нелинейные динамические системы. Это — весьма общая модель, которую потенциально можно использовать в самых разных приложениях, например: ассоциативная память, нелинейная обработка сигналов, моделирование конечных автоматов, идентификация систем, задачи управления.

Нейронные сети с временной задержкой

Перед тем, как описать собственно динамические сети, рассмотрим, как сеть с прямой связью используется для обработки временных рядов. Метод состоит в том, чтобы разбить временной ряд на несколько отрезков и получить таким образом статический образец для подачи на вход многослойной сети с прямой связью. Это осуществляется с помощью так называемой разветвленной линии задержки, см. рис. 1.6.

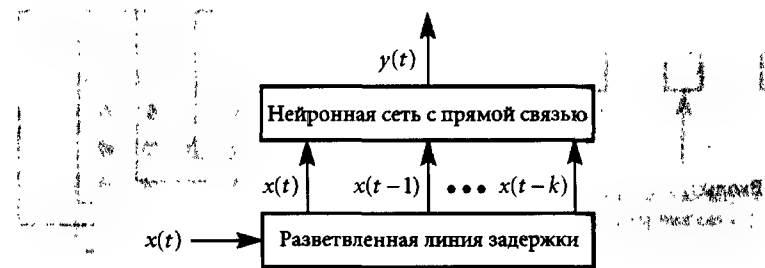


Рис. 1.6. Нейронная сеть с временной задержкой

Архитектура такой нейронной сети с временной задержкой позволяет моделировать любую конечную временную зависимость вида

$$y(t) = F[x(t), x(t-1), \dots, x(t-k)]. \quad (21)$$

Поскольку рекуррентные связи отсутствуют, такая сеть может быть обучена при помощи стандартного алгоритма обратного распространения ошибки или какого-то из его многочисленных вариантов. Сети такой конструкции успешно применялись в задачах распознавания речи, предсказания нелинейных временных рядов и нахождения закономерностей в хаосе.

Сети Хопфилда

С помощью рекуррентных сетей Хопфилда можно обрабатывать неупорядоченные (рукописные буквы), упорядоченные во времени (временные ряды) или пространстве (графики, грамматики) образцы. Рекуррентная нейронная сеть простейшего вида была введена Хопфилдом; она построена из N нейронов, связанных каждый с каждым, причем все нейроны являются выходными.

Сети такой конструкции используются, главным образом, в качестве ассоциативной памяти, а также в задачах нелинейной фильтрации данных и грамматического вывода. Кроме этого, недавно они были применены для предсказания (см. [92], [74]) и для распознавания закономерностей в поведении цен акций (см. [154]).

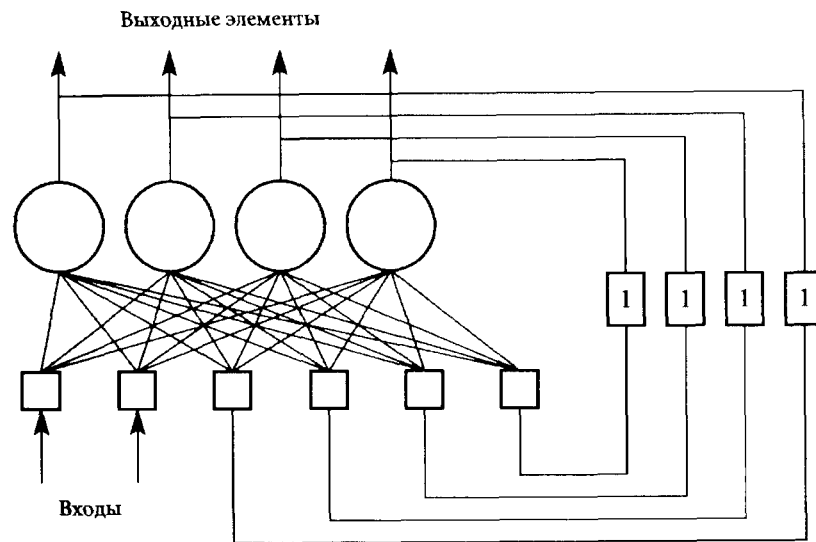


Рис. 1.7. Сеть Хопфилда

Самоорганизующиеся сети

Введенную Кохоненом [167] «самоорганизующуюся карту признаков» можно рассматривать как вариант нейронной сети. Сеть такого типа рассчитана на самостоятельное обучение: во время обучения сообщать ей правильные ответы необязательно. В процессе обучения на вход сети подаются различные образцы. Сеть улавливает особенности их структуры и разделяет образцы на кластеры, а уже обученная сеть относит каждый вновь поступающий пример к одному из кластеров, руководствуясь некоторым критерием «близости».

Сеть состоит из одного входного и одного выходного слоя. Количество элементов в выходном слое непосредственно определяет, сколько различных кластеров сеть сможет распознать. Каждый из выходных элементов получает на вход весь входной вектор. Как и во всякой нейронной сети, каждой связи приписан некоторый *синаптический вес*. В большинстве случаев каждый выходной элемент соединен также со своими соседями. Эти внутрислойные связи играют важную роль в процессе обучения, так как корректировка весов происходит только в окрестности того элемента, который наилучшим образом откликается на очередной вход.

Выходные элементы соревнуются между собой за право вступить в действие и «получить урок». Выигрывает тот из них, чей вектор весов окажется ближе всех к входному вектору в смысле расстояния, определяемого, например, евклидовой метрикой. У элемента-победителя это расстояние будет меньше, чем у всех остальных¹. На текущем шаге обучения менять веса разрешается только элементу-победителю (и, может быть, его непосредственным соседям); веса остальных элементов при этом как бы заморожены. Выигравший элемент изменяет свой весовой вектор, немного перемещая его в сторону входного вектора. После обучения на достаточном количестве примеров совокупность весовых векторов с большой точностью приходит в соответствие со структурой входных примеров — векторы весов в буквальном смысле моделируют распределение входных образцов.

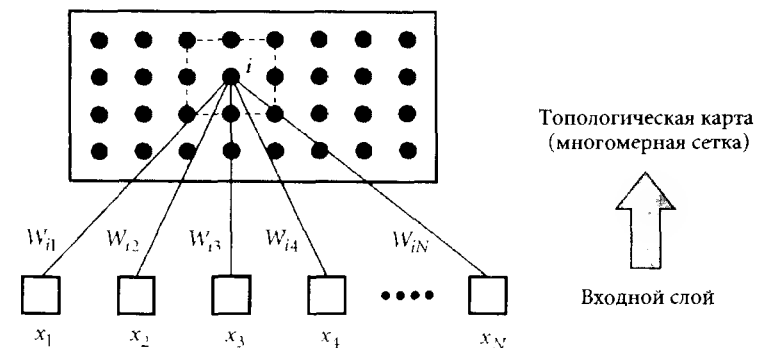


Рис. 1.8. Самоорганизующаяся сеть Кохонена. Изображены только связи, идущие в i й узел. Окрестность узла показана пунктиром

Очевидно, для правильного понимания сетью входного распределения нужно, чтобы каждый элемент сети становился победителем одинаковое число раз — весовые векторы должны быть *равновесными*. Как установил Де Сиено [88], для этого необходимо присутствие некоторого механизма «справедливости». Один из возможных

способов осуществления этого механизма следующий: в расстояние между входным и весовым вектором вносится некоторая добавка, положительная для тех элементов, которые часто выигрывают, и отрицательная для тех, которые чаще проигрывают. Таким образом, шансы проигрывающих повышаются, и вся картина выравнивается. Величина добавки меняется в процессе обучения в соответствии с изменениями частоты выигрывшей.

Перед началом работы сети Кохонена нужно сделать две вещи:

- векторы весов должны быть случайно распределены по единичной сфере,
- все весовые и входные векторы должны быть нормированы на единицу.

Сеть со встречным распространением

Сеть со встречным распространением (CPN, Counterpropagation Network) соединяет в себе свойства самоорганизующейся сети Кохонена и концепцию Outstar-сети Гроссберга. В рамках этой архитектуры элементы слоя сети Кохонена не имеют прямого выхода во внешний мир, а служат входами для выходного слоя, в котором связям адаптивно придаются веса Гроссберга. Эта схема возникла из работ Хехт-Нильсена [138]. CPN-сеть нацелена на постепенное построение искомого отображения входов в выходы на основе примеров действия такого отображения. Сеть хорошо решает задачи, где требуется способность адаптивно строить математическое отображение по его точным значениям в отдельных точках.

Сети данного вида успешно применяются в таких финансовых и экономических приложениях, как рассмотрение заявок на предоставление займов, предсказание трендов цен акций, товаров и курсов обмена валют. Говоря обобщенно, можно ожидать успешного применения CPN-сетей в задачах, где требуется извлекать знания из больших объемов данных.

Обучение CPN-сети складывается из двух процессов адаптации. На первом этапе весовые векторы слоя Кохонена настраиваются так, чтобы моделировать распределение входных векторов. Очевидно, что этот процесс является процессом самостоятельной адаптации. При этом точность аппроксимации будет гарантирована только тогда, когда набор обучающих примеров будет статистически представительным (репрезентативным) для области, на которой действует отображение. Второй адаптационный процесс является несамостоятельным. Он начинается после того, как произошло обучение слоя Кохонена. Происходит настройка весов выходного слоя Гроссберга на примерах с заданным выходом. При этом настраиваются только

веса, соответствующие связям с теми элементами слоя Кохонена, которые активируются (выигравшие элементы, посылающие выходной сигнал «1»), — вектор весов поворачивается в сторону целевого вектора.

Поскольку при подаче на вход очередного вектора может активироваться только один из элементов слоя Кохонена, возможные выходные сигналы CPN-сети совпадают с множеством весовых выходных векторов элементов Кохонена. В случае, когда слой Гроссберга состоит из единственного элемента, получающийся скалярный выход равен одному из весов, соответствующих соединениям этого элемента.

ПРИМЕЧАНИЯ

- ¹ В другом варианте победителем считается элемент, весовой вектор которого имеет наибольшее скалярное произведение с входным вектором. Эта величина также является некоторой мерой близости, потому что скалярное произведение — это проектирование входного вектора на вектор весов. Очевидно, такая проекция будет наибольшей, если векторы имеют близкие направления. При этом методе, однако, оба вектора — весовой и входной — должны быть нормированы по длине, например, быть равными единице. Напротив, евклидово расстояние позволяет работать с векторами произвольной длины.

2

Применение нейронных сетей в задачах классификации и анализа временных рядов

В этой главе архитектура нейронных сетей рассматривается с точки зрения двух наиболее важных видов приложений — задач классификации и анализа временных рядов.

Задача классификации понимается как задача отнесения образца к одному из нескольких попарно непересекающихся множеств. Чаще всего мы будем рассматривать двоичную классификацию. Примерами могут служить определение прибыльности или неприбыльности данной инвестиции, или задача различения жизнеспособных и склонных к банкротству фирм. Задача анализа временных рядов заключается в том, чтобы получить будущие значения некоторой величины, зная ее текущие и прошлые значения и располагая данными о среде.

НЕЙРОННЫЕ СЕТИ В ЗАДАЧАХ КЛАССИФИКАЦИИ

Цель классификации

В предыдущей главе мы рассмотрели методы нейронных вычислений. В настоящей главе мы исследуем две главные области применения сетей с прямой связью: задачи классификации и моделирование временных рядов. Отличие между задачами этих двух типов состоит в наличии (временной) упорядоченности примеров.

Рассмотрим, как нейронные сети с прямой связью (или многослойные перцептроны — MLP, Multilayer Perceptron) используются в задачах классификации. В чем, собственно, заключается эта задача? Во-первых, в любой задаче классификации требуется отнести имеющиеся статические образцы (рукописные буквы, звуковые сигналы, характеристики финансового положения) к определенным классам. Разнообразие примеров, возникающих в реальном мире, практически бесконечно. Эффективность классификации зависит от способа представления этих форм. В числе других здесь имеются следующие способы: распознавание образов, структурное представление и ста-

стистическое представление. В структурном распознавании образов образцы описываются тем, как они составлены из своих компонент, т.е. структурой, подобно тому, как это делается в грамматике языка. Распознавание в этом случае основывается на применении определенных синтаксических правил. При статистическом подходе к распознаванию образец представляется вектором $x \in \mathbb{R}^N$, компоненты которого представляют собой различные характеристики (дескрипторы) образца. Классификатор относит объект x^k к тому или иному классу C в соответствии с определенным разбиением N -мерного пространства, которое называется пространством входов.



Рис. 2.1. Различные степени сложности в представлении классов. Заштрихованные фигуры изображают совокупность всех возможных образцов

При решении задачи распознавания статистическими методами важнейшее значение имеет правильный выбор способа статистического представления объекта. Тем самым, нужно проделать предварительную обработку данных. Для того чтобы выбрать характерные отличительные признаки объектов, требуется, как правило, серьезное изучение исходной проблемы. Например, в моделях банкротства банков важное значение имеют такие показатели, как опыт в управлении фондами и соответствие требованиям адекватности капитала. Различные наборы признаков приводят к разным распределениям. При этом в разных вариантах дисперсия и свойства выпуклости кластеров во входном пространстве могут сильно отличаться, соответственно, при их разделении потребуется проводить границы разной степени сложности — от линейных до сильно нелинейных. Чем лучше была сделана предварительная обработка, тем легче будет решена задача классификации.

Прежде всего нужно определиться с выбором уровня сложности. В реальных ситуациях часто бывает так, что имеется лишь относительно небольшое число образцов, а структура данных позволяет выделить следующие три уровня сложности. Первый (простейший) — когда классы можно разделить прямыми линиями (или гиперплоскостями, если пространство имеет размерность больше двух). Этот случай называется *линейной отделимостью*. Во втором случае одной гиперплоскости для разделения недостаточно (*нелинейная отдели-*

мость), а в третьем случае классы пересекаются, и поэтому разделить их можно только в вероятностном смысле.

В идеальном варианте предварительная обработка должна дать такой набор признаков, чтобы задача оказалась линейно отделимой, — классификация после этого существенно упрощается. К сожалению, это редко удается сделать. Как правило, в нашем распоряжении имеется лишь ограниченный набор образцов, и часть из них используется для проведения границ, разделяющих классы («построение классификатора»). Качество классификатора по отношению к имеющимся примерам измеряется *оценкой*. При последующей работе классификатора с новыми образцами происходит обобщение. Возможные способы оценить способность к обобщению мы рассмотрели в предыдущей главе.

Вероятностная классификация

Вероятность того, что произойдет событие A , обозначается $P\{A\}$. Если, например, событие A состоит в том, что подброшенная монета (правильной формы) упадет вверх орлом, то вероятность $P\{A\}$ равна 0.5. Через $P\{A|B\}$ обозначается *условная* вероятность события A при условии, что произойдет событие B . Вероятность того, что при двух бросаниях монеты оба раза выпадет орел, равна $P\{2A\} = 0.25$. Условная вероятность выпадения двух орлов при условии, что в первый раз выпал орел (событие B), — частный случай условной вероятности, который называется *апостериорной вероятностью*. Так как результаты бросаний монеты независимы, знание первого из них ничего не говорит о втором, и поэтому $P\{A\} = P\{A|B\} = 0.5$. Для задач классификации более характерны зависимые события, когда наши знания о B влияют на ожидаемую вероятность A .

При статистическом распознавании образов оптимальный классификатор относит образец x^k к классу C , руководствуясь решающим правилом Байеса. Для двух классов оно выглядит так:

- Отнести x^k к C_1 , если $P\{C_1|x^k\} > P\{C_2|x^k\}$,
- Отнести x^k к C_2 , если $P\{C_1|x^k\} < P\{C_2|x^k\}$.

Смысл правила простой: образец x^k относится к группе, имеющей наибольшую *апостериорную* вероятность. Это правило оптимально в том смысле, что оно минимизирует среднее число неправильных классификаций. Если имеется такая пара функций $\{\varphi_1(x), \varphi_2(x)\}$, что выполнены условия:

$$\begin{aligned} \varphi_1(x) < \varphi_2(x), & \text{ если } P\{C_1|x\} < P\{C_2|x\}, \\ \varphi_1(x) > \varphi_2(x), & \text{ если } P\{C_1|x\} > P\{C_2|x\}. \end{aligned} \quad (1)$$

то байесовское соотношение между априорной и апостериорной вероятностью сохраняет силу, и поэтому эти функции можно использовать в качестве *упрощенных решающих функций*. Так имеет смысл делать, если эти функции строятся и вычисляются более просто.

Хотя правило выглядит очень простым, применить его на практике оказывается трудно, так как бывают неизвестны апостериорные вероятности (или даже значения упрощенных решающих функций). Их значения можно оценить. В силу теоремы Байеса апостериорные вероятности можно выразить через априорные вероятности и функции плотности по формуле $P\{C_i|x\} = P\{C_i\}P\{x|C_i\} / \sum P\{C_j\}P\{x|C_j\}$, где j — номер класса. Таким образом, правило Байеса для произвольного числа классов принимает вид:

- Отнести x к C_i , если $P\{x|C_i\}P\{C_i\} > P\{x|C_j\}P\{C_j\}$ для всех $j \neq i$.

Классификаторы образцов

Априорную плотность вероятности можно оценить различными способами. В *параметрических методах* предполагается, что плотность вероятности (PDF) является функцией определенного вида с неизвестными параметрами. Например, можно попробовать приблизить PDF при помощи гауссовой функции. Для того чтобы произвести классификацию, нужно предварительно получить оценочные значения для вектора среднего и матрицы ковариаций по каждому из классов данных и затем использовать их в решающем правиле. В результате получится полиномиальное решающее правило, содержащее только квадраты и попарные произведения переменных. Вся описанная процедура называется *квадратичным дискриминантным анализом* (QDA). В предположении, что матрицы ковариаций у всех классов одинаковы, QDA сводится к *линейному дискриминантному анализу* (LDA).

В методах другого типа — *непараметрических* — никаких предварительных предположений о плотности вероятности не требуется. В методе « k ближайших соседей» (k NN) вычисляется расстояние между вновь поступившим образцом и векторами обучающего множества, после чего образец относится к тому классу, к которому принадлежит большинство из k его ближайших соседей. В результате этого границы, разделяющие классы, получаются кусочно линейными. Ряд применений k NN-метода будет рассмотрен в гл. 8. В различных модификациях этого метода используются различные меры расстояния и специальные приемы нахождения соседей. Иногда вместо самого

множества образцов берется совокупность центроидов, соответствующих кластерам в методе адаптивного векторного квантования (LVQ) (см. [168]). В других методах классификатор разбивает данные на группы по схеме дерева. На каждом шаге подгруппа разбивается надвое, и в результате получается иерархическая структура бинарного дерева. Разделяющие границы получаются, как правило, кусочно линейными и соответствуют классам, состоящим из одного или нескольких листьев дерева. Этот метод хорош тем, что он порождает метод классификации, основанный на логических решающих правилах. Идеи древовидных классификаторов применяются в методах построения самонарастающих нейронных классификаторов (см. [194], [244], [50]).

Нейронная сеть с прямой связью как классификатор

Поскольку сети с прямой связью являются универсальным средством аппроксимации функций, с их помощью можно оценить *априорные* вероятности в данной задаче классификации. Благодаря гибкости в построении отображения можно добиться такой точности аппроксимации апостериорных вероятностей, что они практически будут совпадать со значениями, вычисленными по правилу Байеса (так называемые оптимальные процедуры классификации, см. [131]).

Богатые возможности отображения особенно важны в тех случаях, когда на основе нескольких оценок строится высокоуровневая процедура принятия решений. Известно много приложений нейронных сетей с прямой связью к задачам классификации. Как правило, они оказываются эффективнее других методов, потому что нейронная сеть генерирует бесконечное число нелинейных регрессионных моделей (см. [230]).

К сожалению, хотя теоретически характеристики нейронной сети с прямой связью стремятся к байесовской, в применении их к практическим задачам выявляется ряд недостатков. Во-первых, заранее неизвестно, какой сложности (т.е. размера) сеть потребуется для достаточно точной реализации отображения. Эта сложность может оказаться чрезмерно большой. *Архитектура сети*, т.е. число слоев и число элементов в каждом слое, должна быть зафиксирована до начала обучения. Эта архитектура порождает сложные нелинейные разделяющие поверхности в пространстве входов. В сети с одним скрытым слоем векторы образцов сначала преобразуются (нелинейным образом) в новое пространство представлений (пространство скрытого слоя), а затем гиперплоскости, соответствующие выходным узлам, располагаются так, чтобы разделить классы уже в этом новом пространстве. Тем самым, сеть распознает уже другие характерис-

тики — «характеристики характеристик», полученные в скрытом слое.

Все это подчеркивает важность этапа предварительной обработки данных. Чем более компактно представлены характеристики образцов, тем меньше зависимость от настраиваемых параметров сети (0 или 1).

Кодирование на выходе

Задача двоичной классификации может быть решена на сети с одним выходным элементом, который может находиться в состоянии 0 или 1. Для задачи с многими классами нужно разработать *способ записи* (кодирования) *выхода*. Один возможный способ состоит в том, чтобы кодировать k классов с помощью k -мерных наборов, приписывая i -й компоненте значение 1, если исследуемый образец принадлежит i -му классу, и 0 — в противоположном случае. Такой способ часто называют *бабушкиным кодированием*. Другой способ работы с многими классами — разбить задачу с k классами на $k(k-1)$ подзадач, содержащих только по два класса. Окончательное присваивание элементу i -го номера класса осуществляется несложной булевой функцией, на вход которой подаются выходы подзадач. В этом случае число выходных элементов с ростом k растет как k^2 . Это так называемое 2-на-2 кодирование часто оказывается лучше, чем бабушкин метод. Рис. 2.2 иллюстрирует проблему кодирования выхода на примере двумерной задачи с тремя классами. С помощью 2-на-2 кодирования задача классификации решается, тогда как в бабушкином методе кодирования необходимо строить нелинейные разрешающие границы.

Объем сети

Правильный выбор размера сети имеет важное значение. Хорошую, и притом очень маленькую, модель построить просто невозможно, а слишком большая будет чересчур сильно приспосабливаться к обучающим данным и плохо аппроксимировать настоящую задачу. Обычно начинают с сети небольшого размера и постепенно увеличивают ее, пока не будет достигнута нужная точность. При этом обучение сетей на каждом шаге проводится независимо. При другом подходе применяется «алгоритм самонарастания», когда по мере возникновения необходимости в сеть добавляются новые элементы, после чего заново происходит обучение (Stepnet, см. [164]). Упомянем также метод каскадной корреляции (см. [106]). Совершенно другая идея лежит в основе деструктивного подхода: вначале берется сеть завышенного объема и из нее удаляются связи и узлы, существенно не влияющие на решение (см., например, [174]). При этом предполагается, что известна верхняя граница для размера

сети, — чаще всего ее, действительно, можно считать известной. При этом полезно иметь в виду «правило», согласно которому число весов всегда должно быть меньше, чем число обучающих примеров, — иначе сеть будет «привыкать» к данным, т.е. попросту запоминать их, и утрачивать способность к обобщению. Если априорные знания о задаче малы или отсутствуют, объем требуемой сети неизвестен. Например, когда разрешающая граница имеет сферическую форму (один класс внутри другого), число скрытых элементов, необходимых для аппроксимации границы, будет многочленом от N (размерности пространства входов)¹. Для каких-то других задач может потребоваться экспоненциальное число узлов в сети, и наоборот, может оказаться, что задача большой размерности решается с помощью простых границ, составленных из отрезков.

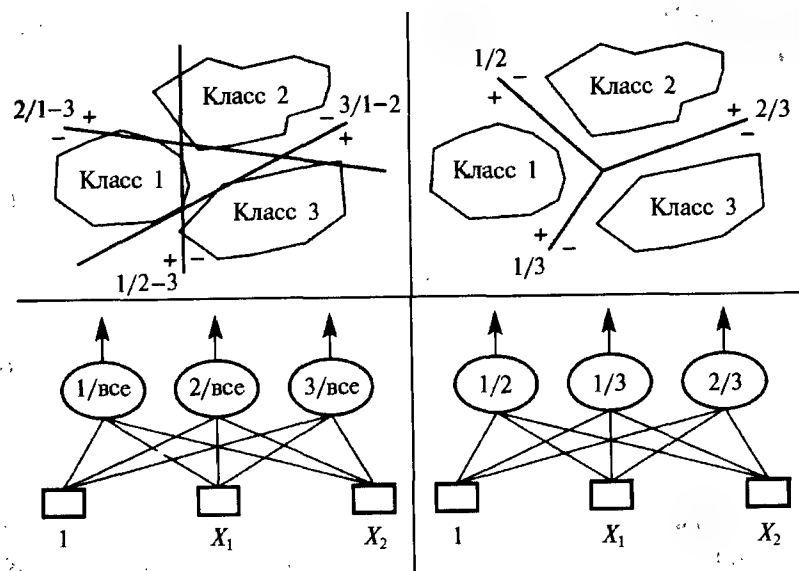


Рис. 2.2. Кодирование выхода на примере двумерной задачи с тремя классами. Слева — «бабушкин» метод кодирования и соответствующая сеть. Справа — 2-на-2 кодирование и соответствующая сеть

Однако правильный выбор объема сети — это еще не все. Надо определить значения всех весов, т.е. сеть должна «научиться» осуществлять нужное отображение. Для этого нужно выбрать эффективный алгоритм обучения. Самое простое здесь — взять классический алгоритм обратного распространения. Однако, часто более эффективными оказываются методы второго порядка. В последнее время было предложено большое количество новых алгоритмов, умень-

шающих время обучения и отбрасывающих субоптимальные решения.

Выбор архитектуры сети

Обычно опробуется несколько конфигураций с различным числом элементов и структурой соединений. Одними из наиболее важных показателей являются объем обучающего множества и обеспечение способности к обобщению при дальнейшей работе, и нужного результата можно достичь на различных схемах. Чаще всего используются процедуры последовательного спуска (с подтверждающим множеством) или N -кратного перекрестного подтверждения. Могут быть применены и более мощные информационные критерии (IC): обобщенное перекрестное подтверждение (GCV), итоговая ошибка предсказания Акаике (FPE), критерии Байеса (BIC) и Акаике (AIC) (см. [103]). Для того чтобы улучшить способности к обобщению и устранить опасность переобучения, применяются также уменьшение весов и их исключение (прореживание дерева). При этом изменяется архитектура сети: удаляются некоторые связи и изучается, какое влияние они оказывали на эффективность.

Анализ показателей работы сети

После того, как выбор модели (т.е. архитектуры сети) сделан и проведена ее проверка, ее можно использовать для предсказания, объяснения и диагностики. С ее помощью можно определять, к какому из классов принадлежит предъявленный образец, или изучать возможные связи между различными характеристиками объектов и принимаемым решением, или выявлять причины, повлекшие за собой неправильную классификацию.

В результате обработки поступившего образца классифицирующая модель выдает на выходе некоторое значение. Как правило, эти выходные значения бывает необходимо подвергнуть доводке. Например, если класс с номером 2 кодируется выходным вектором $(0, 1, 0)$, а сеть выдала $(0.1, 0.6, 0.3)$, то нужно решить, имеются ли достаточные основания причислить объект ко 2-му классу.

«Настройка» классификатора обычно бывает основана на пороговых правилах и/или сравнении расстояний между значениями целевых показателей. Нужно помнить, что нейронная сеть с прямой связью и сигмоидными выходами выдает ответ в непрерывном виде, обычно в интервале от 0 до 1 в зависимости от того, как располагаются разделяющие гиперплоскости скрытых элементов. Однако, даже если на выходе используются не апостериорные вероятности, а какая-либо более простая решающая функция, имеется возможность выдать надежный ответ. Настраивая критерий отбрасывания, можно

уменьшить долю неправильных классификаций за счет повышения доли сомнительных случаев, т.е. таких ситуаций, когда лучше, вообще, не принимать никакого решения, чем принять неправильное.

Хотя на нейронные сети часто смотрят как на «черный ящик», есть некоторые возможности выяснить влияние каждого фактора на решение, принимаемое в задаче классификации. На данное время формального метода, позволяющего извлекать из обученной сети информацию о задаче или о правилах классификации, не существует. Как правило, анализ сетей проводится эвристически (см. [127]). В рамках этой книги мы не имеем возможности рассмотреть все описанные в литературе методы выяснения стратегии классификации, которую осуществляет сеть. С некоторыми из таких методов мы встретимся позже (в частности, в гл. 4) при рассмотрении процедур моделирования.

Далее может быть проведен анализ чувствительности и исследование причин неправильной классификации. Исследуя, насколько сильно (или, наоборот, слабо) сеть реагирует на отдаленность образца от разделяющей границы, можно вывести характеристику «разрешающей способности» метода классификации. Одно из возможных применений такого анализа — исследование сомнительных случаев и последующее удаление их из обучающей базы данных.

По завершении всех указанных процедур сеть можно использовать в сложных комплексах принятия решений в сочетании с традиционными подходами, а также с другими сетями, обученными независимо и настроенными на другие характеристики объектов.

Сводка действий при построении классификатора

В построении классификатора на основе нейронной сети можно выделить следующие этапы:

- 1 Данные:
 - а) Составить базу данных из примеров, характерных для поставленной задачи.
 - б) Разбить всю совокупность данных на два множества: обучающее и тестовое.
- 2 Предварительная обработка:
 - а) Выбрать систему признаков, характерных для данной задачи, и преобразовать данные соответствующим образом для подачи на вход сети (нормировка, стандартизация и т.д.). В результате желательно получить линейно отделимое представление множества образцов.

- б) Выбрать систему кодирования выходного значения или значений («бабушкино» кодирование, 2-на-2 и др.).
- 3 Конструирование, обучение и оценка качества сети:
 - а) Выбрать топологию сети: число элементов и структуру связей (входы, слои, выходы).
 - б) Выбрать функцию активации, которая будет использоваться.
 - в) Выбрать подходящий алгоритм обучения сети.
 - г) Оценить качество работы сети (по подтверждающему множеству или по информационному критерию) в зависимости от ее сложности с целью оптимизировать архитектуру (уменьшение весов, прореживание и т.д.).
 - д) Остановиться на варианте сети с наилучшей способностью к «обобщению» и оценить качество работы по тестовому множеству.
- 4 Использование и диагностика:
 - а) Выяснить степень влияния различных факторов на принимаемое решение (эвристики).
 - б) Убедиться, что доля случаев неправильной классификации достаточно мала.
 - в) При необходимости вернуться к шагу 2, изменив способ представления образцов или предварительно «почистив» базу данных.
 - г) Приступить к практическому использованию сети.

Для того чтобы сконструировать хороший классификатор, очень важно иметь в своем распоряжении высококачественные данные. Никакой метод построения классификаторов, будь то в области распознавания образов, машинного обучения или многомерной статистики, никогда не выдаст классификатор нужного качества, если имеющийся набор примеров не будет достаточно богатым и представительным для той популяции, с которой придется работать в данной модели.

Пример: ирисы Фишера

Завершая наше введение в методы классификации при помощи нейронных сетей, рассмотрим одну задачу распознавания образов, которую часто берут за образец при проверке методов. Это — задача Фишера об ирисах. Мы вкратце перечислим результаты, полученные при помощи классических подходов, а затем сравним их с тем, что дают нейронные сети.

База данных и предварительная обработка

Имеются данные измерений для 150 экземпляров ирисов, в равных частях (по 50 штук) принадлежащих к трем видам (*iris setosa*, *iris versicolor*, *iris virginica*). Для каждого экземпляра ириса известны 4 величины: длина чашелистика (SL), ширина чашелистика (SW), длина лепестка (PL) и ширина лепестка (PW) (см. [111]). На рис. 2.3 показан представительный срез базы данных. Входной файл для нейронной сети состоит из 150 строк (по 50 для каждого сорта):

- первые 4 переменные — длина чашелистика (SL), ширина чашелистика (SW), длина лепестка (PL) и ширина лепестка (PW).
- пятая переменная — целевая, обозначает класс (вид) и для различных видов принимает следующие значения: 0 — *setosa*, 0.5 — *versicolor*, 1 — *virginica*. Такой способ кодировки связан с предположением Фишера, что *versicolor* — это гибрид *setosa* и *virginica*. На срезе PW/PL (рис. 2.3) можно заметить соответствующую упорядоченность. Здесь можно было бы использовать и другой метод кодировки, например разбить задачу на три подзадачи с двумя классами каждая.

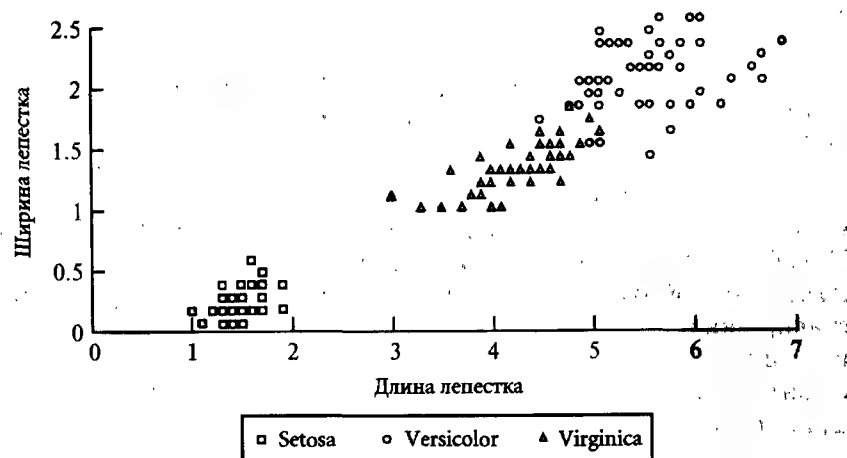


Рис. 2.3. Ирисы Фишера. Двумерная проекция массива данных (ширина и длина лепестка)

Обучение и доводка

На этой задаче были опробованы три различные конфигурации сети. Все сети имели четыре входных узла, количество скрытых элементов менялось от трех до пяти, выходной узел был во всех случаях один. В качестве отображающей функции для каждого узла был взят простой сигмоид, принимающий на выходе значения от 0 до 1. Об-

разец относился к тому или иному классу согласно простому пороговому правилу интерпретации выхода сети:

- выход $> 2/3 \rightarrow virginica$,
- выход $< 1/3 \rightarrow setosa$,
- остальные значения $\rightarrow versicolor$.

Обучение было проделано с помощью пакета Nevada QuickProp на майнфрейм-комьютере Convex. Никакой оптимизации процесса обучения не производилось: процесс заканчивался, когда средняя квадратичная ошибка переставала существенно уменьшаться за 1000 итераций (эпох). Коэффициент, определяющий длину шага, во всех случаях принимался равным 0.1, а начальные значения весов брались случайным образом в интервале $[-0.1, 0.1]$. Обучение занимало 10000–30000 эпох в зависимости от сложности сети. Полученные в результате значения весов приведены в верхнем левом углу таблицы на рис. 2.7 (см. ниже). В частности, значения веса третьего скрытого элемента и выходного узла равны, соответственно, 48.65 и 0.69. Вес соединения, идущего от первого скрытого элемента к выходному, равен 4.16.

Результаты

В табл. 2.1 приведены результаты для сетей в сравнении с другими известными методами.

Следует при этом помнить, что невязка при повторной подстановке дает несколько приукрашенную картину для точности метода. Особенно это относится к самой большой сети, на которой достигнуто полное распознавание.

Метод	Невязка при повторной подстановке
Правило индукции (Вейс и др.)	7/150
Один ближайший сосед	6/150
Деревья классификации и регрессии	4/150
Индукция по решающему дереву	3/150
Линейный дискриминантный анализ	3/150
Квадратичный дискриминантный анализ	3/150
Общий дискриминантный анализ	3/150
Сеть 4-3-1	3/150
Сеть 4-4-1	2/150
Сеть 4-5-1	0/150

Таблица 2.1. Результаты классификации (невязка при повторной подстановке) для задачи с ирисами Фишера

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

Задача анализа временных рядов

Временной ряд — это упорядоченная последовательность вещественных чисел x_t , $t = 1, 2, \dots, T$, представляющих собой результаты наблюдений некоторой величины. Эти значения обычно получают как результаты измерений в некоторой физической системе. Если нас интересуют зависимости между текущими и прошлыми значениями, то нужно рассматривать вектор задержки $(x_{t-1}, x_{t-2}, \dots, x_{t-n})$ в n -мерном пространстве сдвинутых во времени значений, или *пространстве задержки*.

Цель анализа временных рядов состоит в том, чтобы извлечь из данного ряда полезную информацию. Для этого необходимо построить математическую модель явления. Такая модель должна *объяснять* существо процесса, порождающего данные, в частности — *описывать* характер данных (случайные, имеющие тренд, периодические, стационарные и т.п.). После этого можно применять различные методы фильтрации данных (сглаживание, удаление выбросов и др.) с конечной целью — *предсказать* будущие значения.

Таким образом, этот подход основан на предположении, что временной ряд имеет некоторую математическую структуру (которая, например, может быть следствием физической сути явления). Эта структура существует в так называемом *фазовом пространстве*, координаты которого — это независимые переменные, описывающие состояние динамической системы². Поэтому первая задача, с которой придется столкнуться при моделировании — это подходящим образом определить фазовое пространство. Для этого нужно выбрать некоторые характеристики системы в качестве фазовых переменных. После этого уже можно ставить вопрос о предсказании или экстраполяции. Как правило, во временных рядах, полученных в результате измерений, в разной пропорции присутствуют случайные флуктуации и шум. Поэтому качество модели во многом определяется ее способностью аппроксимировать предполагаемую структуру данных, отделяя ее от шума.

Что могут дать в этом отношении нейронные сети? В этой главе будет показано, что нейронные сети можно рассматривать как обобщение традиционных подходов к анализу временных рядов. Нейронные сети дают дополнительные возможности в моделировании нелинейных явлений и распознавании хаотического поведения. Благодаря своей большой гибкости (на одной топологии можно реализовать много различных отображений), сети могут ухватывать самые разные структуры в фазовом пространстве.

Статистический анализ временных рядов

Подробное описание методов статистического анализа временных рядов выходит за рамки этой книги. Мы вкратце рассмотрим традиционные подходы, выделяя при этом обстоятельства, которые имеют прямое отношение к предмету нашего изложения. Начиная с пионерской работы Юла [295], центральное место в статистическом анализе временных рядов заняли линейные модели ARMA. Со временем эта область оформилась в законченную теорию с набором методов — теорию Бокса–Дженкинса (см. [48]). В этом подходе модель задается двумя компонентами, характеризующими *авторегрессию* и *скользящее среднее*. Общая формула для процесса с авторегрессией и скользящим средним порядка (p, q) имеет вид:

$$x_t = a_0 + \sum_{j=1}^p a_j x_{t-j} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t, \quad (2)$$

где p — порядок авторегрессии (положительное целое число), q — порядок скользящего среднего, ε_t — шум (некоррелированный временной ряд, подчиненный гауссову распределению с нулевым средним и дисперсией σ_ε^2). Коэффициенты a_j и b_j являются параметрами модели. Если $q=0$, то получается авторегрессионная модель $AR(p)$, а если $p=0$, — модель скользящего среднего $MA(q)$.

Присутствие в модели ARMA авторегрессионного члена выражает то обстоятельство, что текущие значения переменной зависят от ее прошлых значений. Такие модели называются одномерными. Часто, однако, значения исследуемой целевой переменной связаны с несколькими разными временными рядами. Так будет, например, если целевая переменная — курс обмена валют, а другие участвующие переменные — процентные ставки (в каждой из двух валют). Соответствующие методы называются *многомерными*. Общий вид уравнения многомерной модели такой:

$$x_t = a_0 + \sum_{k=1}^N \sum_{j=1}^{p(k)} a_j^{(k)} x_{t-j}^{(k)} + \varepsilon_t, \quad (3)$$

где k — номер временного ряда (всего их — N). Математическая структура линейных моделей довольно проста, и расчеты по ним могут быть без особых трудностей выполнены с помощью стандартных пакетов численных методов. Следующим шагом в анализе временных рядов стала разработка моделей, способных учитывать нелинейности, присутствующие, как правило, в реальных процессах и системах. Одна из первых таких моделей была предложена Тонгом [261] и называется пороговой авторегрессионной моделью (TAR). В ней, при достижении определенных (установленных заранее) пороговых зна-

чений, происходит переключение с одной линейной AR-модели на другую. Тем самым в системе выделяется несколько режимов работы. Через θ_t обозначим номер режима в момент t ($\theta_t = 1, 2, \dots, r$). Тогда одномерная AR-модель с соответствующим номером дает:

$$x_t = a_0^{(\theta_t)} + \sum_{j=1}^p a_j^{(\theta_t)} x_{t-j}. \quad (4)$$

Затем были предложены STAR-, или «гладкие» TAR-модели. Такая модель представляет собой линейную комбинацию нескольких моделей, взятых с коэффициентами, которые являются непрерывными функциями времени. Примером может служить следующее уравнение модели, в котором θ — гладкая функция, принимающая значения от 0 до 1:

$$x_t = \left(a_0 + \sum_{j=1}^p a_j x_{t-j} \right) [1 - \theta(x_{t-1})] + \left(b_0 + \sum_{j=1}^p b_j x_{t-j} \right) \theta(x_{t-1}) + \varepsilon_t. \quad (5)$$

Были предложены также многочисленные другие нелинейные модели анализа временных рядов, см. [129], [220].

Модели, основанные на нейронных сетях с прямой связью

Любопытно заметить, что все описанные в предыдущем пункте модели могут быть реализованы посредством нейронных сетей. Любая зависимость вида $x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p}) + \varepsilon_t$ с непрерывной нелинейной функцией f может быть воспроизведена на многослойной сети. Пример приведен на рис. 2.4.

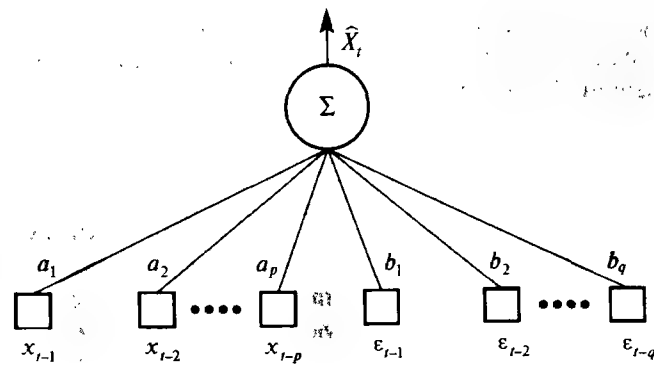


Рис. 2.4. Реализация ARMA (p, q) модели на простейшей нейронной сети

Вместо того, чтобы отображать поверхность во входном (фазовом) пространстве, образованную данными, посредством одной ги-

перплоскости (AR), нескольких гиперплоскостей (TAR), или нескольких гиперплоскостей, гладко соединенных друг с другом (STAR), нейронная сеть может осуществить произвольное нелинейное отображение. Мы говорим это не для того, чтобы представить нейронные сети как универсальную модель в анализе временных рядов, а просто чтобы показать все многообразие структур, которые таким способом можно моделировать. Недавние исследования показали, что нейронные сети имеют, по сравнению с классическими моделями, более высокие потенциальные возможности при анализе сложной динамической структуры, но при этом дают лучшие результаты и на таких известных типах временных рядов, как стационарные, периодические, трендовые и некоторые другие (см. [275]). Мы согласны с мнением Куама [170], что перед окончательным формированием нейронной сети необходимо проделать моделирование на основе модульного подхода с выделением тренда и сезонных колебаний.

Нейронные сети можно также применять для одномерного и многомерного анализа, должным образом сформировав множество независимых входов и зависящих от них выходов. Как правило, модель строится для того, чтобы предсказывать значения временного ряда для одной целевой переменной, однако, в принципе, модель может предсказывать значения и нескольких переменных (например, доходы по акциям на различное время вперед), если в сеть добавить дополнительные выходные элементы.

При этом, однако, исследования в области моделирования временных рядов при помощи сетей продолжают и в настоящее время, и никаких стандартных методов здесь пока не выработано. В нейронной сети многочисленные факторы взаимодействуют весьма сложным образом, и успех здесь пока приносит только эвристический («кустарный») подход. Типичная последовательность действий при решении задачи прогнозирования финансовых показателей с помощью нейронных сетей показана на рис. 2.5.

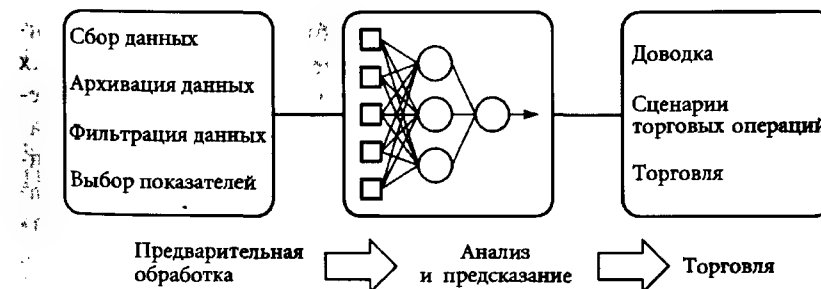


Рис. 2.5. Блок-схема финансового прогнозирования при помощи нейронных сетей

Действия на первом этапе — этапе *предварительной обработки данных*, очевидно, сильно зависят от специфики задачи. Нужно правильно выбрать число и вид показателей, характеризующих процесс, в том числе, — структуру задержек. После этого надо выбрать топологию сети. Если применяются сети с прямой связью, нужно определить число скрытых элементов. Далее, для нахождения параметров модели нужно выбрать критерий ошибки и оптимизирующий (обучающий) алгоритм. Затем, используя средства диагностики, следует проверить различные свойства модели. Наконец, нужно проинтерпретировать выходную информацию сети и, может быть, подать ее на вход какой-то другой системы поддержки принятия решений. Далее мы рассмотрим вопросы, которые приходится решать на этапах предварительной обработки, оптимизации и анализа (доводки) сети.

Предварительная обработка данных

При моделировании реальных процессов «чистые» данные — это редкая роскошь. В силу самой своей природы, реальные данные содержат шумы и бывают неравномерно распределены. Очень часто практик просто собирает данные и подает их на вход модели, надеясь, что все получится. Однако, при сетевом подходе (и это верно здесь даже в большей степени, чем для классического статистического анализа) тщательная *предварительная* обработка данных может сэкономить массу времени и уберечь от многих разочарований. Рассмотрим следующие относящиеся сюда вопросы: сбор данных; их анализ и очистку; их преобразование с целью сделать входную информацию более содержательной и удобной для сети.

Сбор данных

Самое важное решение, которое должен принять аналитик, — это выбор совокупности переменных для описания моделируемого процесса. Чтобы представить себе возможные связи между разными переменными, нужно хорошо понимать существо задачи. В этой связи очень полезно будет побеседовать с опытным специалистом в данной предметной области. Относительно выбранных вами переменных нужно понимать, значимы ли они сами по себе, или же в них всего лишь отражаются другие, действительно, существенные переменные. Проверка на значимость включает в себя кросс-корреляционный анализ. С его помощью можно, например, выявить временную связь типа запаздывания (лаг) между двумя рядами. То, насколько явление может быть описано линейной моделью, проверяется с помощью регрессии по методу наименьших квадратов (OLS). Полученная после оптимизации невязка R^2 может принимать значения от 0 (полное несоответствие) до 1 (точное соответствие). Часто бывает так, что для линейных систем OLS-метод дает такие результа-

ты, которые уже нельзя сколько-нибудь значительно улучшить применением нейронных сетей.

В целом, можно сказать, что предварительная обработка через формирование совокупности переменных и проверку их значимости существенно улучшает качество модели. Если никаких теоретических методов проверки в распоряжении нет, переменные можно выбирать методом проб и ошибок, или с помощью формальных методов типа генетических алгоритмов [69], [70].

Очистка и преобразование базы данных

Стоит начать с того, чтобы изобразить распределение переменной с помощью гистограммы или же рассчитать для него характеристики асимметрии (симметричность распределения) и эксцесса (весомости «хвостов» распределения). В результате будет получена информация о том, насколько распределение данных близко к нормальному. Многие методы моделирования, в том числе, — нейронные сети, дают лучшие результаты на нормализованных данных. Далее, с помощью специальных статистических тестов, например, на расстояние Махаланобиса, можно выявить многомерные выбросы, с которыми затем нужно разобраться на предмет достоверности соответствующих данных. Эти выбросы могут порождаться ошибочными данными или крайними значениями, вследствие чего структура связей между переменными может (а может и не) нарушаться (см. [19]). В некоторых приложениях выбросы могут нести положительную информацию, и их не следует автоматически отбрасывать.

Предварительное, до подачи на вход сети, преобразование данных с помощью стандартных статистических приемов может существенно улучшить как параметры обучения (длительность, сложность), так и работу системы. Например, если входной ряд имеет отчетливый экспоненциальный вид, то после его логарифмирования получится более простой ряд, и если в нем имеются сложные зависимости высоких порядков, обнаружить их теперь будет гораздо легче. Очень часто ненормально распределенные данные предварительно подвергают нелинейному преобразованию: исходный ряд значений переменной преобразуется некоторой функцией, и ряд, полученный на выходе, принимается за новую входную переменную. Типичные способы преобразования — возведение в степень, извлечение корня, взятие обратных величин, экспонент или логарифмов (см. [250]). Нужно проявить осторожность в отношении функций, которые определены не всюду (например, логарифм отрицательных чисел не определен). После этого могут быть применены дополнительные преобразования для изменения формы кривой регрессии. Часто это на порядок уменьшает требования к обучению [284], [251].

Для того чтобы улучшить информационную структуру данных, могут оказаться полезными определенные комбинации переменных — произведения, частные и т.д. Например, когда вы пытаетесь предсказать изменения цен акций по данным о позициях на рынке опционов, отношение числа опционов пут (put options, т.е. опционов на продажу) к числу опционов колл (call options, т.е. опционов на покупку) более информативно, чем оба этих показателя в отдельности. К тому же, с помощью таких промежуточных комбинаций часто можно получить более простую модель, что особенно важно, когда число степеней свободы ограничено.

Наконец, для некоторых функций преобразования, реализованных в выходном узле, возникают проблемы с масштабированием. Сигмоид определен на отрезке $[0, 1]$, поэтому выходную переменную нужно масштабировать так, чтобы она принимала значения в этом интервале. Известно несколько способов масштабирования: сдвиг на константу, пропорциональное изменение значений с новым минимумом и максимумом, центрирование путем вычитания среднего значения, приведение стандартного отклонения к единице, стандартизация (два последних действия вместе). Имеет смысл сделать так, чтобы значения всех входных и выходных величин в сети всегда лежали, например, в интервале $[0, 1]$ (или $[-1, 1]$), — тогда можно будет без риска использовать любые функции преобразования.

Еще одна важная проблема (которая одновременно является основным преимуществом нейронно-сетевых методов) — способность работать с данными качественного характера. Отношения эквивалентности или порядка нужно суметь записать для входа (или выхода) сети. Это можно сделать, вводя искусственные переменные, принимающие значения 1 или 0. Одна прикладная задача с качественными данными рассмотрена в гл. 8.

Построение модели

Значения целевого ряда (это тот ряд, который нужно найти, например, доход по акциям на день вперед) зависят от N факторов, среди которых могут быть комбинации переменных, прошлые значения целевой переменной, закодированные качественные показатели. Эти факторы определяются обычными методами статистики (метод наименьших квадратов, кросс-корреляция и т.д.). Входные и выходные переменные преобразуются (масштабированием, стандартизацией) так, чтобы они принимали значения от 0 до 1 (или от -1 до 1). В результате мы получаем первоначальную модель, которую можно попытаться оптимизировать с помощью нейронной сети.

Главной нерешенной проблемой в области анализа временных рядов с помощью нейронных сетей остается определение топологии сети (или числа степеней свободы в модели). Нужно или прямо ука-

зать размеры сети (число скрытых слоев, скрытых элементов, структуру связей), или настроить модель на имеющиеся данные, применяя конструктивный либо деструктивный подход (например, уменьшение весов).

Оптимизация обучения

Следующая задача — найти параметры (веса) модели. Это делается с помощью алгоритма оптимизации (обучения). Известно несколько таких алгоритмов, в частности, методы обратного распространения и «замораживания». Этот этап может занять продолжительное время и потребует большой технической работы (установка начальных значений весов, выбор критерия останова и др.), однако в конце его мы получим некоторую разумную совокупность весов. Нужно следить за тем, чтобы сеть не запоминала шумы, присутствующие во временных рядах (переобучение). Для этого на протяжении всего процесса оптимизации следует проверять, согласуется ли работа модели на обучающем множестве с соответствующими результатами на подтверждающем множестве.

Статическое и адаптивное обучение

При моделировании финансовых временных рядов вопрос о том, как разбить все имеющиеся данные на обучающее, подтверждающее и тестовое множества, является нетривиальным. Например, если данные, касающиеся биржевого краха, отнести к подтверждающему множеству, это даст искаженные результаты. При статическом подходе обычно поступают так: берут два небольших промежутка времени до и после обучающего множества и из них случайным образом выбирают образцы в подтверждающее множество. На рис. 2.6а показан ряд примеров построения подтверждающего множества. Никогда не будет лишним проверить, насколько изменятся результаты, если множество выбрать иначе.

Изменчивый характер финансовых рынков плохо согласуется с долгосрочными моделями устойчивости. Под действием краткосрочных «модных» тенденций или паники на бирже может существенно измениться реакция людей на те или иные показатели рынка. Чтобы справиться с этой трудностью, были предложены так называемые адаптивные нейронные сети, в которых веса модели непрерывно уточняются с помощью процедуры обучения на все новых временных промежутках. Пример такой модели, учитывающей вновь поступающую информацию, показан на рис. 2.6б. Варфиз и Версино [271], [272] применили эту идею для предсказания изменений ежемесячных индексов промышленных и энергетических компаний. Результаты совпали с тем, что получается по модели Бокса-Дженкинса. Де Гроот [84] занимался задачей прогноза обменных

курсов валют. Модель видоизменялась каждые три месяца, и результаты оказались существенно лучше, чем по методу линейной регрессии.

Отбор и диагностика модели

Проверка свойств модели временного ряда — необходимое условие для надежного предсказания и для понимания природы имеющихся закономерностей. К сожалению, этот вопрос слабо освещен в литературе.

Разности между истинными и оцененными значениями должны подчиняться гауссовскому распределению с нулевым средним. Если оказалось, что распределение имеет слишком тяжелые хвосты или несимметрично, то нужно пересмотреть модель. Среди значений разностей могут выявиться закономерности или последовательные корреляции, тогда необходимо дополнительное обучение или улучшение модели.

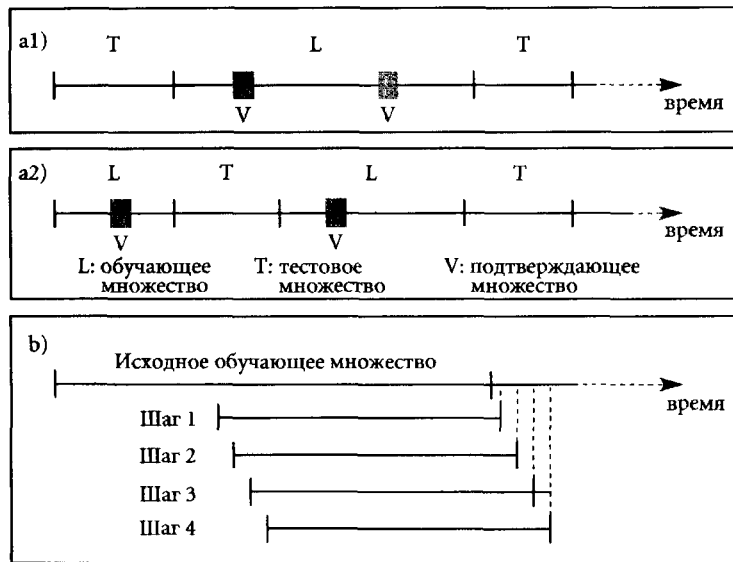


Рис. 2.6. Различные статические методы обучения (а) и обучение по меняющимся промежуткам времени (б)

Оценка качества модели обычно основывается на критерии согласия типа средней квадратичной ошибки (MSE) или квадратного корня из нее (RMSE). Эти критерии показывают, насколько предсказанные значения оказались близки к обучающему, подтверждающему или тестовому множествам. Для рядов с большим разбросом Лапедес [171] предложил критерий средней относительной вариации:

$$\text{arv}(S) = \frac{\sum_{t \in S} (d_t - x_t)^2}{\sum_{t \in S} (d_t - \langle x_t \rangle)^2} = \frac{\sum_{t \in S} e_t^2}{N \sigma^2}, \quad (6)$$

где S — временной ряд, e_t — разность (истинное значение d_t минус x_t) в момент t , $\langle x_t \rangle$ — оценка для среднего значения ряда, N — число данных в ряде. Последующая нормализация с помощью оценки для вариации позволяет проводить более надежные сравнения для различных приложений.

Используя предыдущее значение x_{t-1} целевой переменной, можно оценить способность модели *предсказывать на один шаг вперед*. Подставляя предсказанные значения на место истинных, получим метод *предсказания на k шагов вперед*. Если через несколько шагов модель начинает отклоняться от настоящей траектории, это значит, что в ней присходит накопление и рост ошибки (см. [290], [275]).

Самый распространенный метод выбора нейронно-сетевой модели с наилучшим обобщением — это проверка критерия согласия (MSE, ARV и др.) на тестовом множестве, которое не использовалось при обучении. Если же данных мало, разбивать их на обучающее и подтверждающее множество нужно разными способами. Такое перекрестное подтверждение может потребовать много времени, особенно для нейронных сетей с их длительным процессом обучения.

В линейном анализе временных рядов можно получить несмещенную оценку способности к обобщению, исследуя результаты работы на обучающем множестве (MSE), число свободных параметров (W) и объем обучающего множества (N). Оценки такого типа называются *информационными критериями* (IC) и включают в себя компоненту, соответствующую критерию согласия, и компоненту штрафа, которая учитывает сложность модели. Барроном [30] были предложены следующие информационные критерии: нормализованный IC Акаике (NAIC), нормализованный байесовский IC (NBIC) и итоговая ошибка прогноза (FPE):

$$\begin{aligned} \text{NAIC} &= \ln(\text{MSE}) + \frac{2W}{N}, \\ \text{NBIC} &= \ln(\text{MSE}) + \frac{2W}{N} \ln N, \\ \text{FPE} &= \text{MSE} \left(\frac{1 + W/N}{1 - W/N} \right). \end{aligned} \quad (7)$$

Было показано [198], что FPE представляет собой несмещенную оценку способности к обобщению для нелинейных моделей, в частности, — для нейронных сетей. К сожалению, при этом предполагается, что в нашем распоряжении имеется бесконечное число на-

блюдений, — в этом случае оценка надежности модели, вообще, не представляет особых сложностей.

Ясно, что информационные критерии дают информацию об адекватности модели и помогают выбрать модель подходящего уровня сложности. Другие методы диагностики позволяют, если такая задача стоит, избежать подхода к системе как к «черному ящику». Поскольку основное отличие сети от линейной регрессии — это возможность применять нелинейные преобразователи, имеет смысл посмотреть, насколько глубоко модель использует свои нелинейные возможности. Проще всего это сделать с помощью введенного Вигендом [275] отношения:

$$\mathcal{R} = \frac{\text{вариация сетевых разностей}}{\text{вариация разностей регрессии}} \quad (8)$$

Для более тщательной проверки нелинейных возможностей нужно изобразить распределение выходных значений для скрытых элементов. Слишком большая доля крайних значений (0 или 1) говорит о том, что некоторые элементы попали в режим насыщения³. Еще один способ — построить совместное распределение линейного и нелинейного выходов и применить линейную регрессию. Отклонения от наклона с углом 45° говорят о том, что нелинейные возможности задействованы [84]. Между прочим, встречается точка зрения, что появление во время обучения резких изменений разностей говорит об использовании нелинейностей. К сожалению, это не соответствует действительности.

Следует также проверить, скоррелированы ли действия скрытых элементов. В многомерном регрессионном анализе при росте мультиколлинеарности значения коэффициентов регрессии становятся все менее надежными. Так же и здесь предпочтительно, чтобы выходы скрытых элементов одного слоя были некоррелированы. Нужно найти собственные значения корреляционной матрицы для выходов скрытых узлов по данным обработки всех обучающих примеров. При полной некоррелированности все собственные значения будут равны единице, а отличия от единицы говорят об избыточном числе скрытых элементов. Кроме того, для анализа внутреннего представления нейронно-сетевой модели часто применяются методы кластерного анализа (см. [127]).

Доводка

При построении системы прогноза преследуется цель не только расширить наше понимание процессов, но и получить помощь для принятия решений в финансовой области. Такого рода руководства можно создать с помощью комбинаций нескольких нейронных се-

тей, обученных на разных множествах данных и разных отрезках времени. Например, сигналы на покупку или продажу будут даваться по пороговым значениям, которые настроены с учетом предыдущих позиций и ошибок. Очень важно также распознать момент, когда эффективность модели начинает падать.

Пример: солнечные пятна

Рассмотрим пример применения сетей к анализу классического временного ряда — ряда данных о пятнах на Солнце. Регулярные ежегодные записи этого явления ведутся с 1700 года. Ряд много раз анализировался в статистической литературе, и выяснилось, что он не является ни стационарным, ни линейным, ни гауссовым. Были испробованы различные одномерные методы моделирования временных рядов. Габр и Рао [119] применяли авторегрессионную модель 9-го порядка (с 4 ненулевыми коэффициентами) и билинейную модель. Льюис и Стивенс [179] разработали модель на основе метода многомерных адаптивных регрессионных сплайнов (MARS), а Пристли [221] исследовал модель TAR. В последнее время несколько групп исследователей предприняли попытки проделать анализ ряда с помощью нейронно- сетевого подхода (см. [275], [170], [84]). Результаты, полученные различными методами, собраны в табл. 2.2.

Модель	1700–1920	1921–1955	1956–1979	Число параметров
AR(9)	202	204	286	4
Билинейная модель	124	114	9056	11
MARS	112	142	425	14
TAR	149	129	198	19
Сеть 4-4-1				
ДеГроот	137	141	231	25
Сеть 12-3-1				
Вигенд	127	115	248	43
Сеть 12-3-1				
Куам	129	110	227	43

Таблица 2.2. Средняя квадратичная ошибка, получающаяся в результате анализа различными методами данных о пятнах на солнце

Параметры моделей настраивались по данным за первые 221 год и проверялись на двух последующих периодах (1920–1955 и 1956–1979). Эти два периода отличаются друг от друга наличием выброса,

соответствующего 1956 году, и явной нестационарностью в следующие несколько лет. Очевидно, что авторегрессионные модели оказались слишком примитивны и не дают нужного уровня обобщения. Билинейная и MARS модели, в сравнении с моделью TAR, плохо ухватывают нестационарность во втором тестовом множестве. Нейронные сети различной архитектуры⁴ неплохо показали себя на стационарном тестовом множестве, а на другом — значительно хуже. В целом результаты подтверждают тот факт, что чудес не бывает.

СРАВНИТЕЛЬНАЯ ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ НЕЙРОННЫХ СЕТЕЙ

В идеальном варианте нейронную сеть нужно обучать и применять в моделировании нелинейных систем с постоянной (во времени) структурой и при наличии достаточного объема представительных данных. В этих случаях нейронные сети имеют преимущество перед более простыми методами: экспоненциальным, ARIMA и множественной регрессии.

К сожалению, на практике такие ситуации встречаются нечасто, и, даже если так случайно получилось, требования к надежности решения сводят на нет преимущества модели.

В этом разделе мы исследуем характеристики качества работы нейронных сетей в сравнении с другими методами на примере 18 временных рядов, соответствующих различным показателям экономики Великобритании. Тестовые данные состоят из 18 ежемесячных и 10 ежеквартальных показателей. Все они взяты из базы данных Министерства статистики, имеющейся на базовом компьютере Манчестерского университета.

Предварительная обработка

После загрузки данных в электронную таблицу они были исследованы на присутствие сезонных колебаний (с периодами 1 квартал и 1 год). Там, где сезонные колебания присутствовали, соответствующий показатель брался в качестве одного из входов сети. Входы были линейно масштабированы так, чтобы их значения находились между 0 и 1. Из данных 70 % использовалось в качестве обучающего множества, а оставшиеся 30 % — для оценки.

Обучение

В отсутствие априорной информации о структуре обучение началось с наиболее простой модели: с одним входным, одним скрытым и одним выходным слоем. Далее модель расширялась вплоть до 6-входовой модели с двумя скрытыми слоями, четырьмя узлами в

первом слое и двумя во втором. Во всех случаях вполне хватало одного выходного узла.

Обучение заняло 1000 эпох, причем коэффициенты, определяющие величину шага, на первом уровне полагались равными единице, деленной на число входных узлов, а на втором уровне — вдвое меньше. Использовалась логистическая функция обучения, и результат был лучше, чем для линейной функции, гиперболического тангенса и гауссовой функции.

BCJA	Общее число безработных, тыс.
BFKF	Производство сыра, тонн
BFKA	Поставки молока, тыс. гектолитров
BFKL	Производство сухого молока, тонн
BGAA	Объем импорта, млн. ф. ст.
BFNK	Производство домашнего пива, тыс. гектолитров
BMIB	Туристические поездки, тыс.
BMIA	Суммарная дальность авиарейсов из Великобритании, тыс. км.
APIA	Совокупных доход частных лиц, млн. ф. ст.
BHCD	Производство электроэнергии, квт. ч.
BIAN	Поставки комбикормов фермерам, тыс. тонн
BHIA	Потребление газа, тыс. куб. м.
BHCB	Поставки нефти, тонн
BMLA	Объем пассажирских морских перевозок, тыс.
BMGA	Объем железнодорожных пассажирских перевозок, тыс.
BIFF(S ₁)	Производство формальдегида, тыс. тонн, не SA
BIFF(S ₂)	Производство формальдегида, тыс. тонн, SA
FTAE	Потребление электроэнергии промышленностью, квт.-ч.

Таблица 2.3. Перечень переменных

Результаты работы

Значения, полученные на выходе, преобразовывались обратно в исходный масштаб и анализировались на предмет среднего значения, средней квадратичной ошибки, абсолютной средней ошибки, средней относительной (процентной) ошибки, и показателей Theils μ (Альбург (Ahlburg), 1984).

Прогнозы, которые выдавала сеть, сравнивались с результатами расчетов по другим моделям обработки временных рядов из пакета SPSS PC+. В их числе были различные методы авторегрессии, в том числе методы Холта–Уинтерса и Бокса–Дженкинса. В табл. 2.4 приведены результаты сравнения в терминах Theils μ .

Показатель	Периодичность	Без учета сезонности	С учетом сезонности	Бокс-Дженкинс	Нейронная сеть
BCJA	M	2.64(DN)	1.05(DM)	1.05	1.08
BFKF	M	1.38(DN)	—	0.82	0.18
BFGA	M	—	1.08(DN)	0.68	0.13
BFKL	M	1.86(DN)	1.35(DN)	1.11	0.53
BGAA	M	—	0.35(DM)	0.30	0.15
BFNK	M	1.12(DN)	0.69(DM)	1.24	0.09
BMIB	M	4.37(DN)	0.45(DM)	1.66	0.44
BMIA	M	—	0.56(DM)	1.74	0.37
APIA	Q	0.64(Holt)	—	0.82	1.87
BHCD	Q	—	0.81(LN)	0.91	0.41
BIAN	Q	—	1.30(NM)	2.74	0.24
BHIA	Q	0.17(DN)	—	0.14	0.11
BHCB	Q	1.08(DN)	3.39(DM)	0.96	0.98
BMLA	Q	0.73(DN)	—	0.44	0.40
BMGA	Q	0.84(DN)	0.59(DM)	0.65	0.41
BIFF(S ₁)	Q	1.774(LN)	—	2.06	0.19
BIFF(S ₂)	Q	0.98(LN)	—	0.94	0.17
FTAE	Q	1.09(DN)	0.42(DM)	0.33	0.40

DM — тренд убран, мультипликативная сезонность

DN — тренд убран, сезонность не учитывается

Holt — Холт-Дженкинс, линейная, сезонность не учитывается

LN — линейный тренд, сезонность не учитывается

NM — мультипликативная сезонность, тренд не учитывается

Таблица 2.4. Результаты сравнения

Результаты анализа столь представительного набора рядов различных экономических показателей оказались весьма обнадеживающими: из 10 ежеквартальных показателей в трех случаях нейронные сети продемонстрировали примерно такую же эффективность, как и модель Бокса-Дженкинса, в шести случаях — лучшую, и только в одном — для доходов частных лиц — заметно худшую.

Для семи из восьми рядов ежемесячных показателей сеть дала значительно лучшие результаты, чем модель Бокса-Дженкинса ARIMA, и в одном случае ошибки были одного порядка. Суммируя результаты по всем 18 рядам, можно сказать, что стандартная модель нейронной сети очень хорошо справилась с анализом экономических показателей, имеющих различную периодичность, характер сезонных изменений и выражение (натурное либо денежное). Резуль-

таты работы модели существенно лучше, чем у авторегрессионных моделей и модели Бокса-Дженкинса. Этот эксперимент показал, что значительные затраты времени на построение и обучение нейронной сети вполне себя оправдывают.

Нейронно-сетевые модели можно совершенствовать еще и еще, но даже при относительно простом подходе получается довольно устойчивая архитектура. При обработке реальных данных, с шумом и меняющейся структурой, всегда приходится заботиться о том, чтобы не происходило переобучение, и универсальные (а не сделанные специально для данного ряда) модели дают в этом смысле определенную защиту.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

К настоящему времени разработано много программных пакетов, реализующих нейронные сети. Вот некоторые, наиболее известные программы-симуляторы нейронных сетей, представленные на рынке программного обеспечения: Nestor, Cascade Correlation, Neudisk, Mimenice, Nu Web, Brain, Dana, Neuralworks Professional II Plus, Brain Maker, HNet, Explorer, Explorenet 3000, Neuro Solutions, Prapagator, Matlab Toolbox. Стоит также сказать о симуляторах, свободно распространяемых через университетские серверы (например, SNNS (Штутгарт) или Nevada QuickPropagation). Важным качеством пакета является его совместимость с другими программами, задействованными в обработке данных. Кроме того, важны дружественный интерфейс и производительность, которая может доходить до многих мегафлопсов (млн. операций с плавающей точкой в секунду). Платы-ускорители позволяют сократить время обучения при работе на обычных персональных компьютерах, однако для получения надежных результатов с помощью нейронных сетей, как правило, требуется мощный компьютер.

В качестве примера рассмотрим простую сеть, моделирующую описанную выше задачу Фишера про ирисы с помощью электронной таблицы Excel. Как уже говорилось, в верхнем левом углу таблицы на рис. 2.7 расположены веса.

Последовательность действий при моделировании:

- ввести в клетку I13 формулу $\leq \text{Sigmoid}(\$B4\$ + \text{SumProduct}(C13:F13, \$C4\$:\$F4\$)) >$, где *Sigmoid* — имя определяемого вами макроса для вычисления стандартного сигмоида, а *SumProduct* — скалярное произведение двух массивов одинаковой размерности (в данном случае — вектора весов W_j и входного вектора X_k). Обратите внимание на отдельное слагаемое — пороговый коэффициент $\$B4\$$,

Банкротства, паники и безумия

В этой главе мы рассмотрим следующий вопрос: обладают ли финансовые рынки внутренним механизмом нелинейной обратной связи. Если такой механизм, внешне проявляющийся в якобы случайном, хаотическом поведении цен, действительно, существует, то это бросает серьезный вызов таким известным и широко принятым теориям, как теория *случайного блуждания* и *гипотеза эффективного рынка*. Мы возьмем несколько простых и хорошо известных моделей, основанных на предположении о хаотическом поведении, сгенерируем с их помощью временные ряды и внимательно изучим каждый из них. Затем на этих временных рядах мы проведем ряд экспериментов с использованием нейронных сетей. Это позволит нам выяснить, насколько нейронные сети способны обнаруживать (и предсказывать) детерминированные закономерности, на основе которых ряды были получены. Там, где это возможно, мы будем сравнивать качество прогноза, выдаваемого нейронной сетью, с тем, что дает модель линейной регрессии.

ТЕОРИЯ ХАОСА И РЫНКИ КАПИТАЛА

В своей часто цитируемой книге «Хаос и порядок на рынке капитала» [214] Петерс говорит, что в рамках теории хаоса можно получить ряд моделей, способных воспроизводить развитие событий на финансовых рынках. В этих моделях предлагается детерминированное объяснение для некоторых аспектов поведения финансовых цен, которые, вообще говоря, считаются случайными и непредсказуемыми.

Тот факт, что хаотические модели дают хорошее приближение для финансовых временных рядов, говорит о важности изучения поведения рынков капитала как нелинейных динамических процессов и является дополнительным доводом в пользу применения в задачах прогноза различных нелинейных методов, в том числе, — нейронных сетей. В любом случае применение нейронных сетей согласуется с принципом ограниченной рациональности Саймона [243], согласно которому эффективность рынка ограничена в силу ограниченных возможностей человека в обработке информации.

Есть и другие, более практические причины роста интереса к изучению нелинейных характеристик финансовых рынков. В последние

годы рынок претерпел определенную структурную перестройку: торговля вышла за рамки одного государства и может происходить в любой точке земли; система комиссионных цен и затрат на совершение сделок пришла в беспорядок; время, затрачиваемое на совершение сделок, существенно уменьшилось; благодаря производным финансовым инструментам (таким, как, например, опционы) снизились требования к резервированию капитала при проведении арбитражных операций.

Если брать долгосрочный аспект, то представляется вполне правдоподобным, что решающее влияние на цены рынка оказывают такие экзогенные факторы, как курсы обмена валют, процентные ставки, показатели экономического роста, тенденции издержек и прибыли. В такой ситуации рынок удовлетворительно описывается так называемой «гипотезой эффективного рынка» (см. [107]), согласно которой вся поступающая информация учитывается в текущих ценах рынка, и будущие изменения цен зависят только от будущей информации. Напротив, для более коротких сроков могут быть разработаны более продвинутые методы прогноза, использующие обратные связи и различные технические и структурные факторы.

Устоявшиеся парадигмы финансовой науки, такие, как модель случайного блуждания и гипотеза эффективного рынка, предполагают, что финансовые рынки реагируют на информацию рационально и плавно. В этом случае едва ли можно придумать что-то лучше линейных связей и стационарного (обрабатывающего тренды) поведения.



Рис. 3.1. Влияние информации различного типа на изменения цен

К сожалению, в реальном поведении финансовых рынков мы видим не просто обращение трендов, но постоянно возникающие несоответствия курсов, волатильность, явно не отвечающую поступающей информации, и периодически случающиеся скачки уровня цен и волатильности. Для описания поведения финансовых рынков были разработаны и имели определенный успех некоторые новые моде-

ли хаотического поведения. В этой главе в качестве стандартного набора данных, на котором будут сравниваться нейронные сети и модели линейной регрессии, мы будем использовать выходную информацию некоторых моделей хаоса.

Эффективность рынка определяется эмпирически, и улучшенные модели выявляют различные отклонения, касающиеся малых фирм, фирм с повышенным уровнем прибыли, а также эффектов конца рабочей недели. Такие нелинейные модели, как множественный дискриминантный анализ (MDA), выявляют задержки в реакции рынка, даже если речь идет об обычных, открыто публикуемых годовых отчетах (см. [9]).

БАНКРОТСТВА, ПАНИКИ И БЕЗУМИЯ

На коротких отрезках времени, когда рынок живет без (или с небольшим количеством) вновь поступающей экзогенной информации, представляет интерес влияние таких эндогенных факторов, как неидеальное распространение информации, рост курсов из-за больших объемов покупок (или падение из-за больших продаж), ликвидность рынка, нелинейность механизмов корректировки курсов.

Если считать, что изменения цен, вопреки соображениям эффективности на продолжительных отрезках времени, определяются многочисленными и часто нелинейными обратными связями, то на основе теории хаоса можно построить улучшенные модели, описывающие «влияние прошлого на настоящее» (см. [213]–[216]). Драматические обвалы рынка при отсутствии существенных изменений информации, резкие изменения условий доступа и сроков при пересечении компанией какого-то невидимого порога в кредитной сфере — все это проявления нелинейности. Реальное поведение финансовых рынков, скорее, противоречит правилам обращения линейных трендов, чем подтверждает их.

Рассмотрим простую модель финансового рынка, в которой для продавцов привлекательна высокая, а для покупателей — низкая рыночная цена. Если при достаточно низких ценах покупатели испытывают трудности с финансированием, то характер обратной связи будет нелинейным. Этот эффект можно учесть, возводя уровень цены в степень, большую единицы, например, в квадрат. Если коэффициенты обратной связи для продавца обозначить f^s , а для покупателя — f^b , то сама обратная связь для покупателя и продавца запишется так:

$$p_t = f_0^b + f_1^b p_{t-1} + f_2^b p_{t-1}^2, \quad (1)$$

$$p_t = f_0^s + f_1^s p_{t-1} + f_2^s p_{t-1}^2, \quad (2)$$

где $f_0^b > 0$, $f_1^b < 1$ и $f_0^s < 0$, $f_1^s > 1$.

Из (1) и (2) получаем:

$$p_t = c_0 + c_1 p_{t-1} + c_2 p_{t-1}^2, \quad (3)$$

где $c_0 = 0.5(f_0^b + f_0^s)$, $c_1 = 0.5(f_1^b + f_1^s)$ и $c_2 = 0.5(f_2^b + f_2^s)$.

Если $c_0 = 0.5$, $c_1 = -0.7$, $c_2 = 1.2$, то по значениям цены (предполагается, что они лежат в интервале от 0 до 1) в предыдущий момент времени можно вычислить текущую цену. На рис. 3.2 показана так называемая логистическая кривая¹ (или логистическое отображение), получающаяся при выбранных значениях параметров. Для цен ниже 0.4 видна сильная положительная обратная связь, тогда как для цен выше 0.4 имеется менее сильная отрицательная обратная связь.

Здесь имеются две точки устойчивого рыночного равновесия (т.е. когда $p_t = p_{t-1}$) — они соответствуют точкам пересечения кривой с прямой линией, идущей под углом 45° из начала координат. Это — точки 0.42 и 1. Можно также заметить, что при начальном значении цены, равном нулю, в следующий момент она подскочит до 0.5.

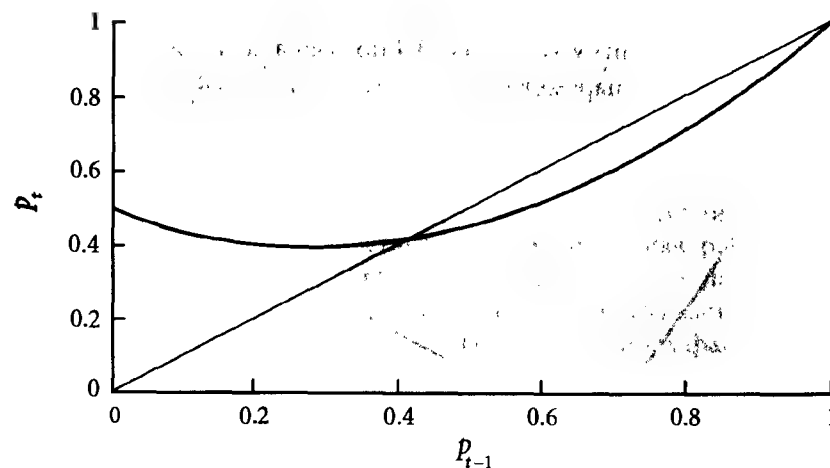


Рис. 3.2. Логистическая кривая; квадратичная обратная связь

Если вводить в систему случайные возмущения цены (т.е. ее начальные значения в интервале от 0 до 1), то, в отсутствие других влияний, через некоторое число шагов система постепенно придет в состояние равновесия. Как видно из рис. 3.3, независимо от начального возмущения, равновесное значение цены равно 0.42.

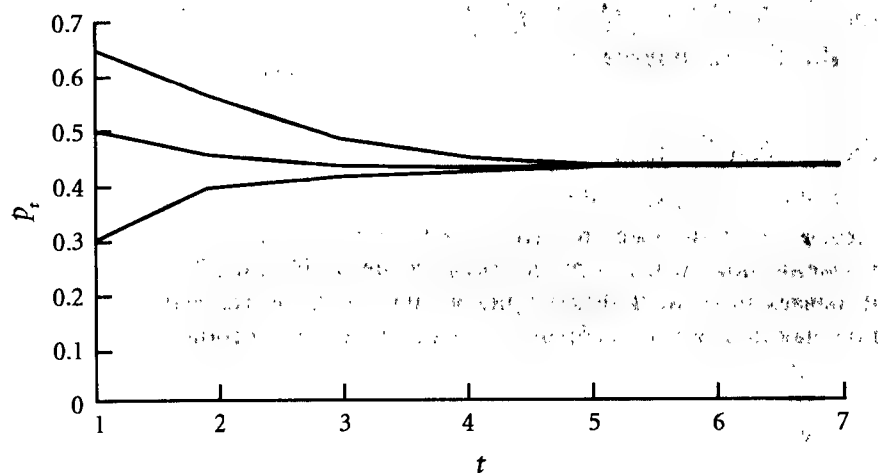
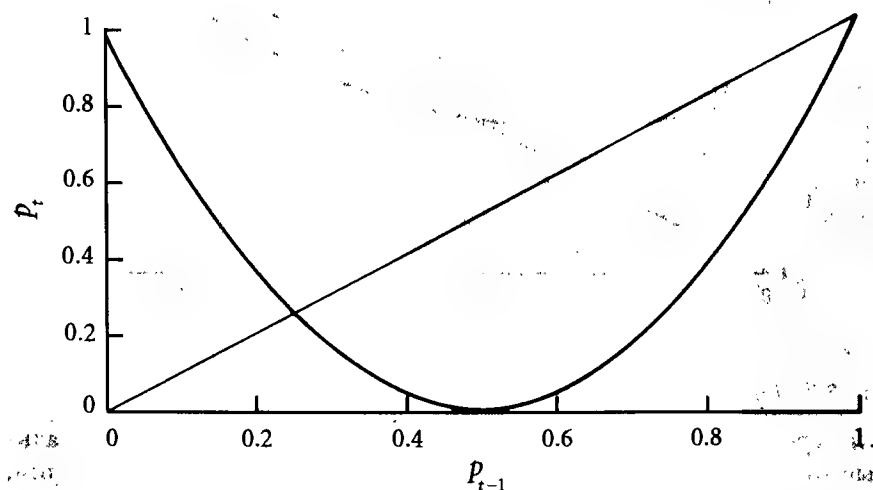


Рис. 3.3. Точечный аттрактор простейшей нелинейной системы

Такое изолированное положение равновесия нелинейной системы называется *точечным аттрактором*. Разумеется, если цена постоянно будет испытывать случайные изменения, точное равновесие никогда не будет достигнуто. На рис. 3.4 показана логистическая кривая для другого набора параметров: $c_0 = 1$, $c_1 = -4$, $c_2 = 4$.

Рис. 3.4. Вид логистической кривой при $c_0 = 1$, $c_1 = -4$, $c_2 = 4$

В этом случае система будет обладать следующим свойством: какое бы начальное значение в интервале от 0 до 1 ни было взято, полученная в результате последовательность цен будет не сходящейся, а будет испытывать внешне случайные колебания без видимой пе-

риодичности и какого-либо окончательного предельного значения (рис. 3.5). Системы такого типа подробно описаны в литературе по хаотическим процессам¹.

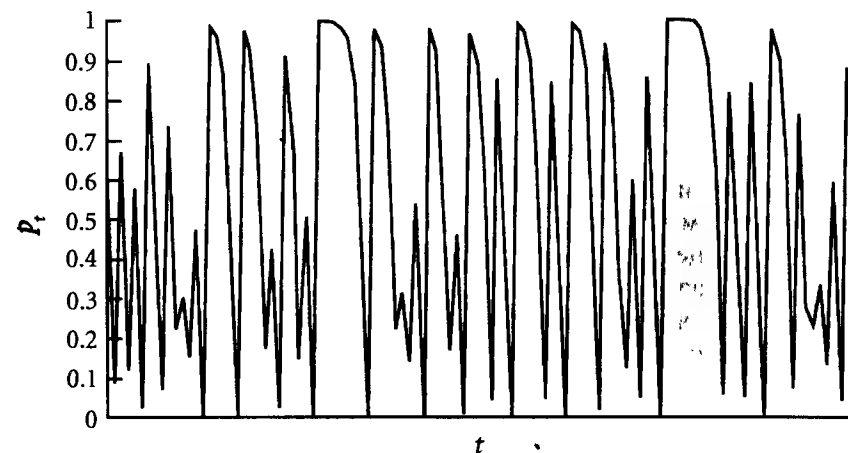


Рис. 3.5. Внешне случайное поведение цены согласно логистическому отображению

Полученный хаотический ряд интересен тем, что, в отличие от случайного ряда, где значение P_t , в принципе, нельзя определить, зная P_{t-1} , здесь все подчинено точным и детерминированным соотношениям. Более того, изменения цены в этой модели вызваны не «случайно» поступающей новой информацией, а нелинейностью самого рынка! Конечно, в действительности динамика рынка будет сложнее, чем здесь представлено, и адекватная модель должна учитывать также то, какое влияние на цены рынка оказывает вновь поступающая информация (в виде основных экзогенных факторов или в какой-то иной форме). Но и в этом случае логистическая модель образования цены может выдавать более или менее хаотические колебания цен.

Если система не стремится к положению равновесия, а все время совершает беспорядочные колебания (как в случае логистической модели с параметрами рис. 3.4), то говорят, что система имеет *странный аттрактор* (в отличие от рассмотренного ранее точечного аттрактора). Помимо этих двух крайних случаев, возможна ситуация, когда система совершает периодические движения вокруг некоторого равновесного состояния. У системы, таким образом, имеется целое семейство решений. Такое поведение называется *предельным циклом*. Для нашей логистической модели эта ситуация возникнет, если значения параметров c_1 и c_2 будут лежать в интервале от 2.5 до 3.75.

МОЖНО ЛИ ПРЕДСКАЗЫВАТЬ ЗАКОНОМЕРНОСТИ ВО ВРЕМЕННОМ РЯДЕ ЦЕН?

Какие выводы позволяет сделать нелинейная динамика в отношении справедливости гипотезы эффективного рынка? Действительно ли в ценах рынка учитывается вся относящаяся к делу информация, а колебания цен отражают случайный характер поступающих данных, или же в основе этих случайных колебаний лежат предсказуемое рассеяние информации и нелинейность динамики рынка? Возможно, связи между поступающей информацией и последующими изменениями цен не такие уж неопределенные, как это принято считать, и доля случайности в изменениях цен преувеличена. Эти выводы открывают возможность для (хотя бы частичного) предсказания будущих изменений цены.

Из сказанного следует также тот вывод, что, обнаружив новую закономерность, мы, вообще говоря, не сможем исключить ее влияние. Так, в нашей простой модели саморегулирующегося рынка причиной нелинейной обратной связи был ограниченный капитал. Если все участники рынка испытывают такие структурные ограничения, то они просто не смогут применять прибыльную тактику торговли, основанную на улучшенном прогнозе. Таким образом, закономерность во временном ряду цен сохранится, хотя при этом рынок будет эффективным в техническом смысле этого понятия.

Характер динамики рынка связан с его формальными и неформальными установками, институциональным окружением, а также сложившейся практикой, предвзятыми мнениями, тенденциями и психологическими барьерами. Все это делает возможным существование (до поры скрытых и выглядящих как случайности) закономерностей в поведении рыночных цен, и эти закономерности, в принципе, можно предсказывать. Научные исследования выявляют все новые и новые закономерности такого рода. Интересный пример приведен в работе [86]. Ее авторы заметили, что рынок иностранных валют не любит спот-курсы, выраженные круглыми числами, например, курс обмена фунт стерлингов/доллар ровно в 1.5000. Исследование реальных рыночных курсов однозначно подтверждает эту закономерность, однако использовать эту «аномалию» непросто, так как даже если мы будем знать, что сделка по цене 1.5000 сулит прибыль, нам трудно будет найти партнера по сделке.

Перед тем, как обратиться к описанию экспериментов с нейронными сетями в распознавании скрытых закономерностей, мы хотели бы отметить, что знание природы хаотического процесса не позволяет существенно улучшить качество долгосрочных (многопериодных) прогнозов. Причина этого в том, что даже сравнительно не-

большие ошибки прогноза, полученные на первом этапе, будут в дальнейшем расти и приводить на последующих итерациях ко все большему расхождению истинного и прогнозируемого рядов.

Это явление ухудшения прогноза можно увидеть на рис. 3.6. Здесь в качестве приближения к «истинному» начальному значению 0.4 было взято значение 0.41. На первых шести шагах ряды ведут себя сходным образом, но затем начинают расходиться все больше и, в конце концов, становятся совершенно некоррелированными, так что близость их значений может быть лишь случайной.

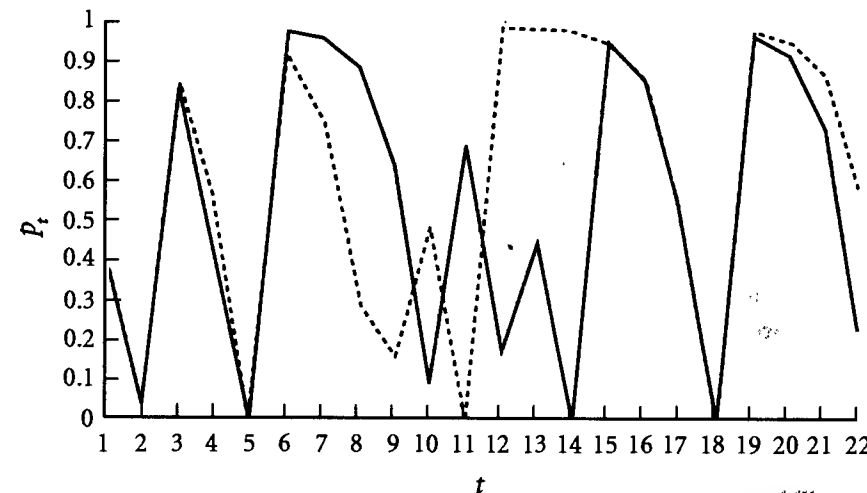


Рис. 3.6. Различное поведение временных рядов при почти одинаковых начальных значениях

НЕСКОЛЬКО НЕЙРОННО-СЕТЕВЫХ ЭКСПЕРИМЕНТОВ С ЛОГИСТИЧЕСКИМИ ВРЕМЕННЫМИ РЯДАМИ

Идет ли речь об эффективности рынка или о каких-либо других закономерностях, в любом случае выявить регулярное поведение во внешне случайной последовательности цен — очень интересная задача для любого аналитика. В примере временного ряда цен, показанном на рис. 3.5, от исследователя потребуется определенная квалификация, чтобы суметь выделить совершенно четкую, хотя и скрытую, детерминированную закономерность. Савиту [237] принадлежит замечание, что структуру хаотических систем легче понять, рассматривая их в более высоких размерностях. К сожалению, такое многомерное пространство может быть выбрано бесконечным числом способов. Стандартный критерий Бокса-Дженкина автокорре-

ляции первого порядка предполагает линейность системы и не дает ключа к пониманию связей между p_{t-1} и p_t , видимых на диаграмме. То же самое относится и к таким более сложным методам анализа временных рядов, как ARIMA². Самое большое, что можно будет получить, — это то, что очень высокие и очень низкие цены (и, соответственно, большие прибыли и потери) встречаются часто. Таким образом, распределение доходов будет иметь тяжелые хвосты и будет сильно отличаться от нормального. Статистические методы, которыми чаще всего пользуются экономисты, являются линейными и плохо отражают нелинейные динамические свойства саморегулирующихся рынков³. Конечно, можно воспользоваться линейными методами с предварительным преобразованием, например, вместо какой-то из переменных взять ее логарифм или более высокую степень, но здесь возникают проблемы с выбором из слишком большого количества возможных преобразований. Кроме того, методы, основанные на средней квадратичной ошибке, плохо справляются с разрывами в траекториях хаотической модели.

Представляется, что нейронные сети лучше, чем другие методы, подходят для выявления нелинейных закономерностей в отсутствие *априорных* знаний об основной модели. Их можно применять во всех случаях, где обычно используются линейные (или преобразованные линейные) методы с проведением оценок посредством подходящего статистического метода (см. [114]). Чтобы лучше представить себе возможности нейронных сетей, рассмотрим процесс обучения очень простой многослойной сети с алгоритмом обратного распространения ошибки (MBPN) на искусственно смоделированном псевдохаотическом временном ряде. Начнем с простого логистического механизма обратной связи, описанного выше (см. рис. 3.4):

$$p_t = 1 - 4p_{t-1} + 4p_{t-1}^2. \quad (4)$$

Рассматривалась 1-2-1 MBPN-сеть (1 входной элемент, 1 скрытый слой из двух элементов и 1 выходной элемент), с полной связью между соседними слоями и без прямых связей между входом и выходом. Сеть обучалась на последовательности входных значений p_{t-1} и выходных p_t , полученных из соотношения (4). Мы использовали обобщенное нормальное обучающее правило дельта с коэффициентом обучения 0.9. Задача оказалась довольно простой. Для периода обучения было взято 3000 циклов, после чего мы предсказывали значения цены на 100 периодов вперед, а затем сравнивали их с «настоящими» значениями, вычисленными по формуле. Результаты представлены на рис. 3.7.

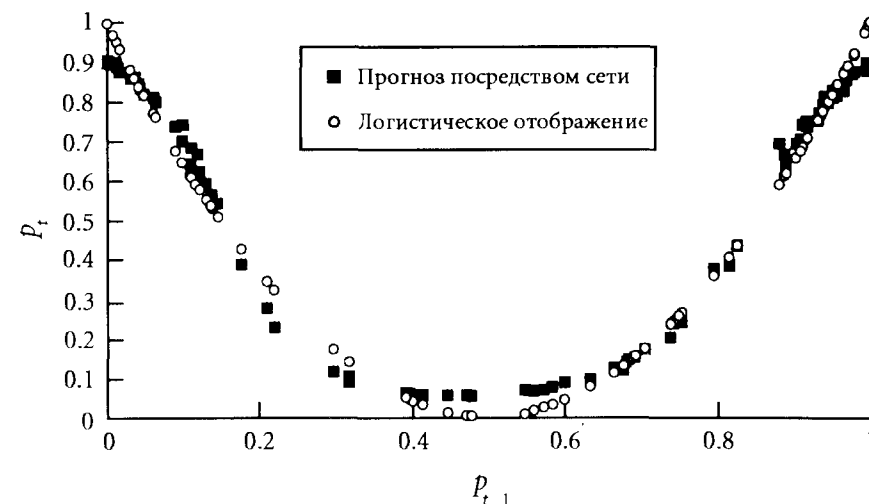


Рис. 3.7. Прогноз логистического отображения с помощью 1-2-1 сети

В области очень высоких и очень низких цен имеются некоторые отклонения, но в целом результаты выглядят вполне удовлетворительно.⁴ Корреляция между целевой переменной и выходом сети составляет 0.9898, а среднеквадратичная ошибка прогноза равна 0.0026. При более длительном обучении и другом выборе параметров обучения можно добиться большей точности.

Вернемся теперь к первому варианту системы с квадратичной обратной связью (см. рис. 3.2) с параметрами $c_0 = 0.5$, $c_1 = -0.7$, $c_2 = 1.2$, но теперь добавим к ней белый шум. Предположим, что изменения цены наполовину зависят от неинформированных инвесторов, которые реагируют на текущую цену в соответствии с приведенным выше соотношением, а на другую половину — от реакции информированных инвесторов на поступающую информацию о рынке, которая носит случайный характер. В нашей модели эта вторая составляющая цены бралась из нормального распределения⁵ с нулевым средним и стандартным отклонением 0.2. Результирующий доход есть среднее арифметическое от случайного дохода и дохода, определяемого обратной связью.

Жирной линией на рис. 3.8 изображены цены с учетом всех видов дохода, а тонкой линией — цены, полученные на основе предсказанного дохода без учета вновь поступающей (случайной) информации. Хорошо видно сглаживающее действие обратной связи: низкая цена дает сильную положительную обратную связь, а высокая — небольшую отрицательную. В табл. 3.1 приведены данные линейной регрессии между последовательными значениями цены p_{t-1} и p_t для варианта с суммарными доходами. Поскольку сдвиг близок к нулю, а ко-

эффицент для p_{t-1} приблизительно равен единице, мы не можем отвергнуть гипотезу случайного блуждания, в соответствии с которой ожидаемое значение для p_t равно p_{t-1} .



Рис. 3.8. Искусственно смоделированные ряды, описывающие информированных и неинформированных инвесторов

	Коэффициенты	Стандартная ошибка	t статистика	P-значение	Нижние 95%	Верхние 95%
Сдвиг	0.035425	0.020927	1.692821	0.093666	-0.00611	0.076959
p_{t-1}	0.924581	0.040076	23.07093	2.16E-41	0.845042	1.00412

Таблица 3.1. Регрессия последовательных значений цены

На рис. 3.9 представлена зависимость p_t от p_{t-1} для ряда с полным учетом доходов. Как видно из рисунка, линия регрессии и квадратичная обратная связь вносят свой вклад в изменения цен. На этой точечной диаграмме трудно заметить нелинейную связь, потому что включенный во временной ряд случайный шум намного интенсивнее исходной детерминированной структуры.

Этот пример является хорошей проверкой способности нейронной сети выявлять исходную структуру. Здесь мы опять использовали полносвязную 1-2-1 сеть без непосредственных связей входа с выходом, которая обучалась с помощью входных значений p_{t-1} и целевых значений p_t . Для обработки сетью многократно подавались наборы из 100 пар значений цен. Как и в первом эксперименте, коэффициент обучения был взят равным 0.9. Во время обучения сети по окончании очередной эпохи (т.е. каждые 100 циклов) вычислялась среднеквадратичная ошибка (MSE). С самого начала этот показатель плавно уменьшался с каждой новой эпохой. Мы продолжали обучение до тех пор, пока MSE не установилась на своем минимальном

значении. Это произошло примерно через 4000 эпох. Затем с помощью обученной сети мы сделали прогноз величины p_t по 100 новым тестовым значениям p_{t-1} . Результаты представлены на рис. 3.10.

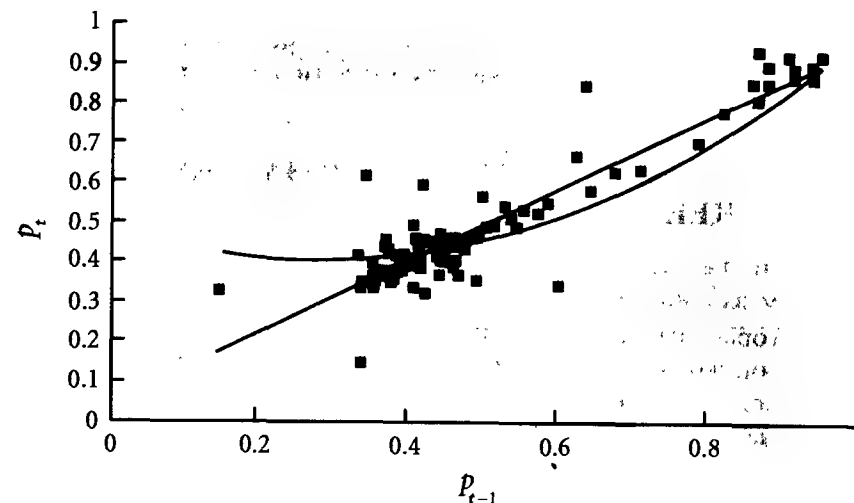


Рис. 3.9. Диаграмма для пар последовательных значений цены, полученная по смоделированному ряду

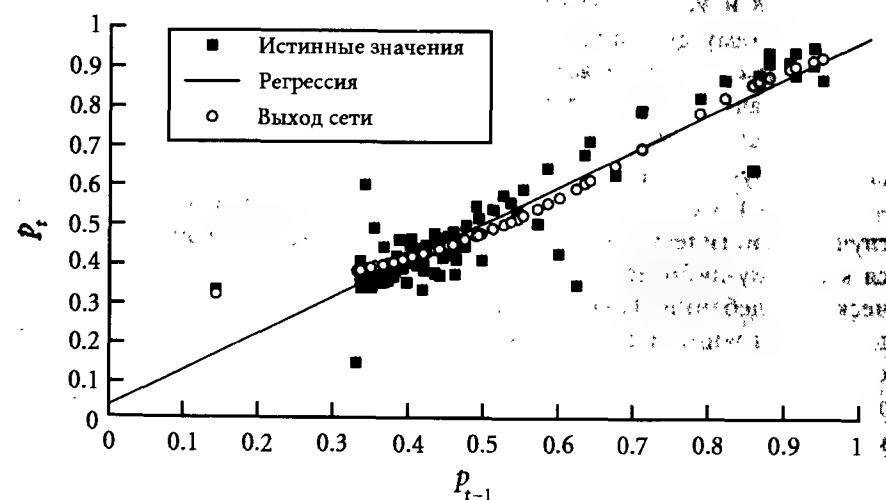


Рис. 3.10. Прогноз методом линейной регрессии и с помощью нейронной сети

Наиболее критическим является участок относительно низких цен, где механизм обратной связи наиболее силен. На диаграмме в числе прочего показаны выход нейронной сети и прогноз по регрессии для таких критических значений p_{t-1} . Очевидно, что сеть вполне удовлетворительно распознает исходную нелинейную взаимосвязь. Интересно при этом, что RMSE (квадратный корень из средней квад-

ратичной ошибки) прогноза нейронной сети (0.0635) всего на 5% лучше, чем у регрессии. Основной вклад в это улучшение вносит именно повышенная точность в области низких цен, где наиболее силен эффект обратной связи. На финансовых рынках время от времени случаются неожиданные события, например, девальвации валют, и здесь кроются наибольшие возможности для извлечения прибыли.

СЕТЕВАЯ ОЦЕНКА В ДВУМЕРНОЙ ЗАДАЧЕ (ОТОБРАЖЕНИЕ ХЕНОНА)

До сих пор в наших экспериментах рассматривались только задачи с одним входным переменным p_{t-1} . Теперь мы обратимся к проблеме, которая аналогична только что рассмотренному одномерному логистическому отображению, но, в отличие от него, имеет двумерный вход. Впервые эта модель была рассмотрена Хеноном [139] и получила название отображения Хенона. Уравнения модели таковы:

$$x_t = 1 + y_{t-1} - ax_{t-1}^2, \quad (5)$$

$$y_t = bx_{t-1}. \quad (6)$$

Как x_t , так и y_t зависят от предыдущих значений x_{t-1} и y_{t-1} , и это делает систему динамической. Из-за квадратичного члена в первом уравнении система является нелинейной. Если мы возьмем произвольные начальные значения и сгенерируем по этим уравнениям ряд значений для x_t и y_t , то окажется, что их значения беспорядочно и внешне случайно располагаются, соответственно, в интервалах от -0.4 до 0.4 и от -1.4 до 1.4 . Так же, как и в рассмотренном ранее случае логистического отображения рис. 3.4, эти значения не сходятся к какому-либо положению равновесия и не совершают периодических колебаний. Таким образом, мы имеем дело с системой, обладающей странным аттрактором. Понятно, что с помощью традиционных статистических методов нам вряд ли удастся выявить структуру модели, поскольку x и y ведут себя беспорядочно (см. [214, с. 152]).

Целью эксперимента должен быть прогноз значения x_t по x_{t-1} и y_{t-1} . Сначала давайте сделаем вид, что мы вообще ничего не знаем о существовании какой-то модели, описывающей ряд x_t , а знаем только (со слов «эксперта»), что здесь играет роль предыдущее значение x_{t-1} , а также еще некоторый показатель y_{t-1} . Естественно начать с линейной регрессии. Чтобы в дальнейшем было удобнее сравнивать регрессию и сеть, промасштабируем значения x и y так, чтобы они лежали на отрезке $[0, 1]$. Полученный в результате ряд для x_t показан на рис. 3.11.

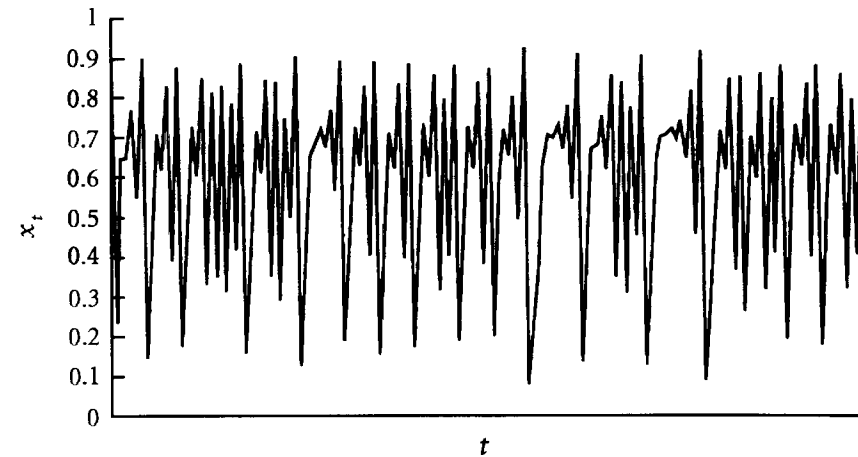


Рис. 3.11. Странный аттрактор отображения Хенона

Чтобы получить исходный материал для последующих экспериментов с нейронными сетями, мы сначала выполнили линейную регрессию на первых 153 членах временного ряда. Результаты регрессии для x_t представлены в табл. 3.2. Коэффициенты регрессии, в том числе сдвиг, существенно отличны от нуля на 95-процентном уровне. Уточненный R^2 равен 0.11.⁶

	Коэффициенты s	Стандартная ошибка	t-статистика	P-значение	Нижний 95%-й уровень	Верхний 95%-й уровень
Сдвиг	0.645396204	0.083048726	7.771295652	1.08143E-12	0.48129985	0.809492557
x_{t-1}	-0.267477915	0.080662827	-3.315999752	0.001142148	-0.426859961	-0.10809587
y_{t-1}	0.18113141	0.089324561	2.027789531	0.044328614	0.004634605	0.357628214

Таблица 3.2. Регрессия для отображения Хенона

Для проверки регрессионной модели мы сформировали прогноз для последних 153 записей в нашей базе данных. Квадратный корень из среднеквадратичной ошибки прогноза регрессионной модели был равен 0.2112. После этого мы обучили 2-2-1 MBPN-сеть на первых 153 совокупностях двух входных и целевой переменных, а вторые 153 записи использовали как подтверждающее множество. Коэффициент обучения, по-прежнему, брался равным 0.9. Обучение прекращалось, если в течение 100 эпох подряд среднеквадратичная ошибка оставалась очень низкой. После этого прогноз был сделан также для 153 образцов.

На рис. 3.12 показаны диаграммы распределения значений x (фазовый портрет) по отношению к значениям y на предыдущем шаге для истинного отображения Хенона, линейной регрессии и MBPN-сети. Квадратный корень среднеквадратичной ошибки нейронной

сети на образцах, не входивших в обучающее множество, составил 0.0281, что существенно ниже, чем соответствующая ошибка регрессии 0.2112. Представляется, что, в отличие от регрессии, сеть довольно хорошо уловила сложную структуру фазового портрета. Это отчетливо видно на рис. 3.12. Хорошие показатели сети станут еще виднее, если мы вычислим истинное и прогнозируемое сетью относительные изменения (R) величины y за один шаг. На рис. 3.13 изображено совместное распределение этих двух величин.

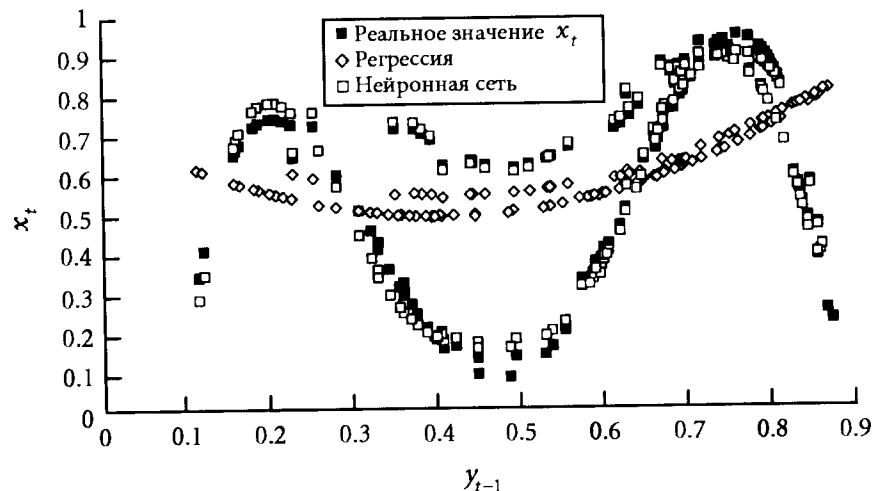


Рис. 3.12. Диаграмма распределения двух последовательных значений ряда Хенона (действительный и прогнозируемый варианты)

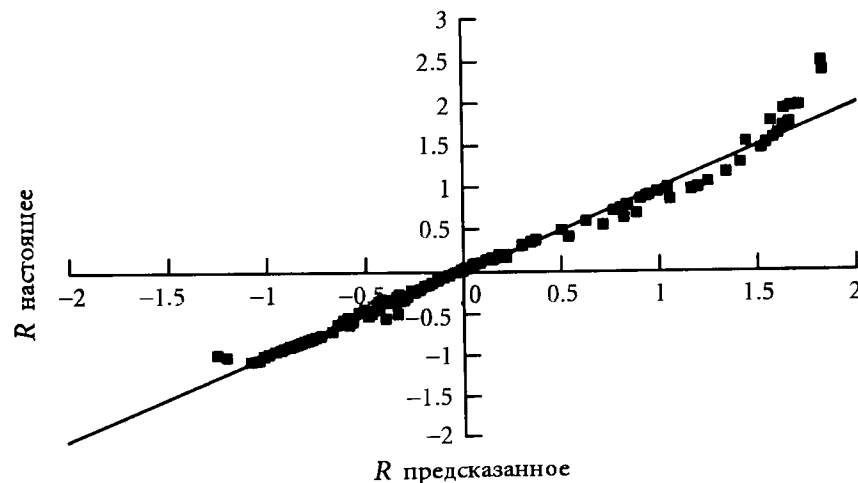


Рис. 3.13. Реальные и прогнозируемые сдвиги по отображению Хенона

Близкое прилегание к прямой, идущей под углом 45° , — очень хороший результат. Для крайних низких и высоких значений сохраняется расхождение, но за счет более длительного обучения и более тщательного выбора архитектуры и параметров сети можно добиться более точной аппроксимации сигмоидальной формы незашумленного процесса (обратите внимание, что «крайних» положительных значений больше). Веса обученной сети показаны на рис. 3.14.

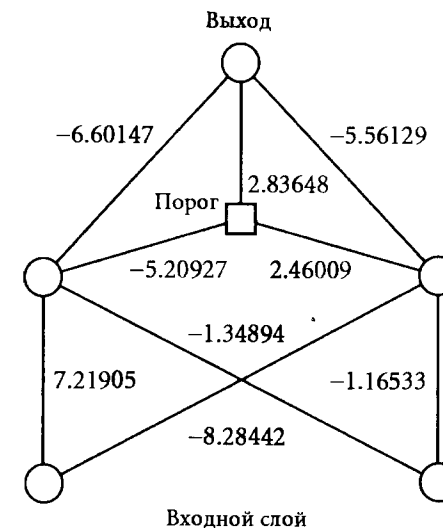


Рис. 3.14. Веса сети с алгоритмом спуска после обучения на отображении Хенона

Левый нижний входной узел соответствует x , а правый нижний — y . Отдельно указаны веса двух скрытых и выходного элемента.

Последний эксперимент с сетью, который мы опишем в этой главе, относится к ряду Хенона с шумом. Мы видоизменили модель следующим образом:

$$\begin{aligned} x_t &= 1 + y_{t-1} - 1.4x_{t-1}^2, \\ y_t &= 0.15x_t + 0.5\epsilon_t. \end{aligned} \quad (7)$$

Случайная составляющая ϵ_t бралась равномерно распределенной на интервале от -0.1 до 0.1 . Таким образом, изменение цены y_t наполовину определяется величиной $0.3x_{t-1}$, отражающей связь с предыдущим моментом, а наполовину — случайной величиной. Исходя из произвольно взятых начальных значений x_{t-1} и y_{t-1} , мы вычислили 306 последовательных значений. Все значения переменных были перемасштабированы так, чтобы они лежали в интервале от 0 до 1. Первые 153 набора использовались для оценки по регрессионной

модели и для обучения нейронной сети, а другие 153 остались для тестирования. Наша задача, по-прежнему, состояла в прогнозировании ряда x_t по значениям x_{t-1} и y_{t-1} .

В табл. 3.3 представлены результаты, касающиеся регрессии. Коэффициент при y_{t-1} очень близок к нулю при 95-процентном доверительном уровне, в то время как сдвиг и коэффициент при x_{t-1} существенно отличны от нуля. Уточненный R^2 равен 0.39. Затем с помощью этой модели был сделан прогноз на 153 шага, при этом квадратный корень из среднеквадратичной ошибки оказался равным 0.1796. Из рассмотрения рис. 3.15 становится ясно, что природа этой ошибки — та же, что была на фазовом портрете рис. 3.12: регрессия не ухватывает существо динамической модели. Очевидно, ошибка метода регрессии недопустимо велика.

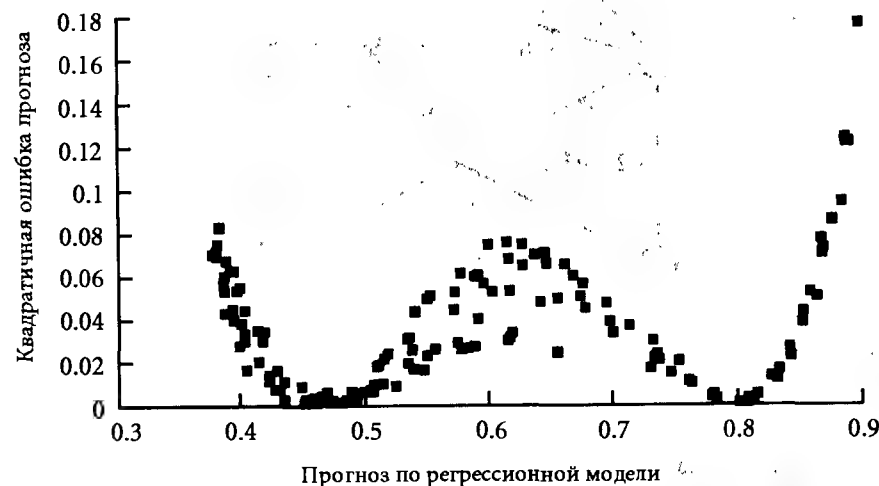


Рис. 3.15. Ошибка регрессионной модели

	Коэффициенты s	Стандартная ошибка	t-статистика	P-значение	Нижний 95% уровень	Верхний 95% уровень
Сдвиг	0.961963	0.078818	12.20484	2.59E-24	0.806226	1.1177
x_{t-1}	-0.63959	0.074895	-8.53976	1.3E-14	-0.78757	-0.4916
y_{t-1}	-0.00654	0.085095	-0.07616	0.939394	-0.17628	0.163198

Таблица 3.3. Регрессия для отображения Хенона с шумом

Применяя к тому же набору данных нейронно-сетевую модель, мы обучали 2-2-1 MBPN-сеть, имеющую те же параметры, что и в случае задачи Хенона без шума. Прodelав обучение из 100 эпох на первой половине данных, мы сделали прогноз относительно другой половины. Чтобы сравнить результаты для сети и регрессии, мы так-

же изобразили на диаграмме (рис. 3.16) совместное распределение прогноза сети и квадратичной ошибки прогноза.

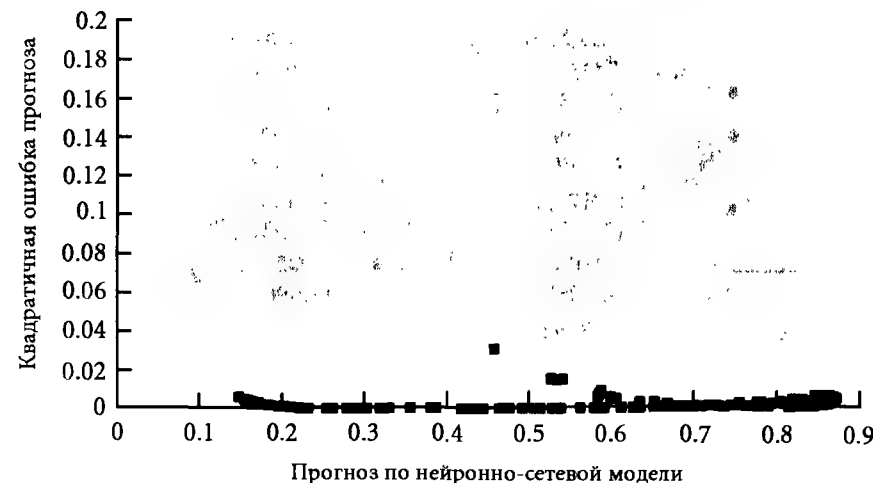


Рис. 3.16. Квадратичная ошибка прогноза многослойной сетью со спуском

В отличие от регрессионной модели, прогноз сети почти не имеет искажений. Соответственно, и RMSE сетевого прогноза (0.0225) значительно меньше, чем у регрессии (0.1796). Можно сделать вывод, что даже в присутствии шума сеть способна распознавать структуру процесса и выдавать надежный прогноз.

УПРОЩЕННЫЙ ВАРИАНТ МОДЕЛИ ХЕНОНА

В силу того, что модель Хенона — двумерная, нейронные сети имеют здесь «фору» перед одномерными методами типа регрессии. Чтобы устранить эту несправедливость, подставим в уравнение (7) выражение для y_t , в результате чего получится следующее уравнение:

$$x_t = 1 + 0.3x_{t-2} - 1.4x_{t-1}^2. \quad (8)$$

Теперь мы можем проделать анализ нашей модели стандартными одномерными методами анализа временных рядов, например, методом Бокса-Дженкинса, а затем сравнить результаты с тем, что дает нейронная сеть с единственным входом, на который подаются предыдущие значения переменной. Как и в предыдущем примере, к процессу Хенона мы добавили 10-процентный случайный шум. Временно представим себе, что мы не располагаем никакой информацией, кроме самих числовых данных. Обычно в таких случаях, начертив данные на графике, пытаются применить модель ARIMA, т.е. стараются найти закономерности типа авторегрессии или скользящего среднего.⁷ В табл. 3.4 представлены результаты анализа методом Бокса-Дженкинса для 5 лагов.

Лаг	Автокорреляция	Частная автокорреляция
1	-0.662 (-11.513)	-0.662 (-11.513)
2	0.425 (5.393)	-0.0216 (-0.0376)
3	-0.359 (-4.170)	-0.152 (-2.643)
4	0.212 (2.332)	-0.145 (-2.522)
5	-0.0745 (0.805)	0.0630 (1.096)
Среднее	0.586927 (2.3283)	

Таблица 3.4. Автокорреляция на обучающих данных для ряда Хенона с шумом

Поскольку после четырех лагов (за исключением 2-го) коэффициенты автокорреляции и частной автокорреляции невелики, мы делаем вывод, что метод AR(4) подходит для этого ряда. Мы делаем прогноз для второй половины набора данных, исходя из четырех предыдущих значений x . Квадратный корень из среднеквадратичной ошибки (RMSE) прогноза равен 0.3642.

Этот же ряд мы проанализировали с помощью многослойной сети (MBPN). Зная вид модели, следовало бы взять сеть с двумя входами, но чтобы сохранить аналогию с четырьмя предыдущими значениями в методе Бокса-Дженкинса, мы выбрали архитектуру 4-2-1. Процесс обучения сети сходиллся не так хорошо, как в предыдущих двумерных примерах. При этом сходимость улучшалась, когда выбирались маленькие начальные значения весов (случайным образом на отрезке $[-0.1, 0.1]$) и коэффициента обучения (0.1). Обучение прекращалось после 4000 эпох. В этот момент RMSE на пробном отрезке ряда (том же самом, что и в методе AR(4)) был равен 0.0649 и продолжал уменьшаться. Различие в точности прогноза сетью и линейной моделью оказалось здесь примерно таким же, как и в предыдущих экспериментах. Это хорошо видно на рис. 3.17, где результаты прогноза по обеим моделям сравниваются с точными значениями.

Как и следовало ожидать (учитывая, что ряд порожден моделью Хенона), прогноз по модели AR(4) имеет искажение параболической формы. При внимательном изучении прогноза сети также можно заметить легкую синусоидальную волну вокруг истинных значений. Однако, амплитуда искажения здесь во много раз меньше.

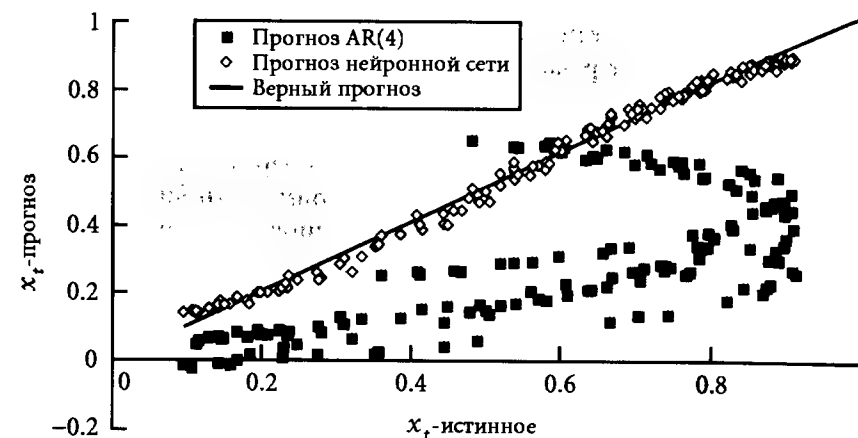


Рис. 3.17. Прогноз по модели AR(4) и по многослойной сети 4-2-1 MBPN

Взяв 4-мерный вход и не предполагая никаких знаний о модели, мы умышленно имели дело с сетью, содержащей лишние элементы. Как это обстоятельство повлияло на формирование весов в результате обучения сети? Для лучшего понимания вопроса в табл. 3.5 веса обученной сети представлены в сравнении с аналогичными весами двумерной сети (последние указаны в скобках). Веса входов, соответствующих $t-1$ и $t-2$, имеют для 4-лаговой и 2-лаговой моделей один и тот же знак и примерно одинаковую величину (по отношению к порогу). Этот факт, а также малая среднеквадратичная ошибка прогноза 4-2-1 сети на новых образцах говорят об устойчивости характеристик сети при добавлении в нее лишних элементов.

Узел	1 ($t-4$)	2 ($t-3$)	3 ($t-2$)	4 ($t-1$)	5 скрытый	6 скрытый	порог
5	0.0335819	0.0880925	0.320498 (0.35698)	-5.11489 (-5.84001)			4.46777 (5.19349)
6	0.090955	0.404738	0.764702 (-1.02628)	-7.35403 (-10.7995)			1.78196 (2.66209)
7 Выход					10.3205 (9.10849)	-5.84991 (-3.86574)	-7.05268 (-6.4816)

Таблица 3.5. Веса обученной 4-2-1 (2-2-1) сети для временного ряда Хенона с шумом

Сравнение результатов прогноза

В заключение этой главы мы сравним качество прогноза, выдаваемого нелинейной нейронной сетью, с тем, что получается по методу линейной регрессии, с помощью другой меры отклонения — Я. Эта мера была предложена Вигендом [275] и определяется как отношение остаточной дисперсии нелинейной модели к остаточной дис-

персии линейной модели. Если две модели работают одинаково хорошо, это отношение равно 1. Малые значения $\mathcal{R}$ указывают на относительное превосходство нелинейной модели. В эксперименте с временным рядом Хенона величина $\mathcal{R}$ была равна 0.12, что говорит о значительном превосходстве MBPN модели перед методом ARIMA. В последующих главах мы будем использовать отношение $\mathcal{R}$ для оценки прогнозов, касающихся реальных рыночных данных.

НЕКОТОРЫЕ ИТоговые замечания

Все описанные в этой главе эксперименты показали хорошую способность MBPN-моделей к обнаружению нелинейных связей во временных рядах финансовых показателей. Это проявлялось в робастности прогноза на тестовых данных. Для сравнения мы применяли также традиционные линейные методы, предполагая при этом, что ничего не знаем о структуре входного ряда. Конечно, имеются более сильные статистические методы, например, такие, где учитывается зависимость дисперсии от прошлых значений (ARCH), или пороговые авторегрессионные модели (TAR), и с их помощью можно находить сложные нелинейные связи. В этой главе мы хотели подчеркнуть тот факт, что методы нейронных сетей не предполагают никаких предварительных знаний о модели. Единственное, что нужно — это значения переменных, а далее сеть уже сама приспосабливается к имеющейся структуре.

Мы хорошо понимаем, что прогнозы, «не зависящие от модели» и сделанные без понимания экономики, — вещь опасная. Мы приветствовали бы любую попытку приоткрыть «черный ящик» нейронной сети. Для этого обязательно потребуются статистические методы — как основные, так и более специальные. В этом смысле нейронные сети можно считать новым полем для приложения этих методов. Основываясь на результатах экспериментов этой главы, мы считаем, что разумным первым шагом для выявления «истинной» модели, описывающей данные, было бы тщательное изучение (с помощью всех доступных средств) прогноза, выдаваемого «наивной», но хорошо обученной MBPN-сетью, и его сравнение с прогнозом по «наивной» линейной модели.

И последнее замечание. Теория хаоса — это лишь одна из теорий, предназначенных для описания экономических явлений. Из того, что поведение финансовых временных рядов может быть воспроизведено, еще не следует, что механизмы нелинейной обратной связи, действительно, существуют. Опытных данных тут пока не так много. Если нелинейные обратные связи в экономической информации полностью отсутствуют, то обнаружить их никакими методами не удастся, и поэтому необходимым первым шагом для анализа

должно стать выяснение факторов, определяющих эндогенное поведение.

ПРИМЕЧАНИЯ

- ¹ В нашей модели парабола смотрит ветвями вверх. Во многих публикациях описаны примеры логистических отображений, где также наблюдается парабола, но она смотрит вниз. Хаотический характер обратной связи в обоих случаях сходный.
- ² Вероятно, для этой задачи лучше подходит не модель ARIMA, а модели семейства ARCH (autoregressive conditional heteroschedasticity). Рассмотрение этих моделей выходит за рамки данной книги. Подробное их описание можно найти, например, в [224].
- ³ Предположение, что доходы на капитал имеют нормальное распределение, в академической литературе часто подвергается сомнению (см. [252]). При этом речь идет о попытках объяснения банкротств и/или эйфорических ценовых шоков.
- ⁴ Здесь мы имеем суперпозицию сигмоидальных функций.
- ⁵ Во многих проведенных нами имитационных тестах мы пользовались случайными выборками из нормального распределения. Такие выборки были получены на основе центральной предельной теоремы математической статистики. Эта теорема утверждает, что сумма независимых равномерно распределенных случайных величин является случайной величиной, распределенной нормально. Сгенерировав k равномерно распределенных случайных величин, мы получаем первую нормальную распределенную величину из формулы:

$$n = \frac{\sum_{i=1}^k \text{RAND}_i - k/2}{\sqrt{k/12}}.$$

Из следующих k равномерно распределенных величин получается вторая нормальная величина, и т.д. При увеличении k эта формула все лучше аппроксимирует нормальное распределение. В пределе $k \rightarrow \infty$ распределение величины n стремится к нормальному распределению с нулевым средним и стандартным отклонением, равным единице. Затем мы преобразуем n в тестовое число по формуле:

$$\text{Тестовое число} = \sigma \cdot n + \mu.$$

Во всех тестах мы брали упрощенный вариант алгоритма с $k = 12$:

$$n = \sum_{i=1}^{12} \text{RAND}_i - 6.$$

Аппроксимация получалась более грубой в области хвостов распределения, т.е. там, где значения переменной наиболее далеки от среднего. Это могло привести к некоторым искажениям в наших тестах. Мы, однако, игнорируем это обстоятельство, поскольку тесты нам нужны только для того, чтобы проиллюстрировать возможности нейронных сетей.

- ⁶ Разумеется, упорный статистик применил бы еще другие критерии, рассмотрел другие спецификации модели и преобразования данных и добился бы лучшего соответствия результатов. В этой главе мы только хотим показать, что обучение MBPN-сети и сравнение сделанного ею прогноза с наивной линейной моделью может служить полезным дополнением к традиционному статистическому подходу.
- ⁷ Применив модель ARIMA и используя полученные по ней результаты, можно затем применить модель ARCH. Здесь мы опять хотим сказать, что нейронная сеть, во всяком случае, может быть полезным добавлением к чисто статистическим методам.

Прогнозирование денежных потоков. Налоговые поступления

Денежные потоки в любой организации, без преувеличения, можно назвать ее кровеносной системой. В то же время этот показатель, как никакой другой, труден для прогнозирования. Эта глава посвящена проблеме управления активами и пассивами Министерства финансов Голландии (далее — MoF). Особое внимание будет уделено оценке суммы ежемесячного валового сбора налогов. Мы рассмотрим и сравним различные методы, в том числе, и модель ARIMA — собственную разработку MoF. Так как нейронные сети превосходят другие методы по показателю среднеквадратичной ошибки (MSE) на вновь предъявляемых образцах, мы будем выделять различные типы индивидуального и совместного поведения переменных с помощью анализа первичных весов, тестов на чувствительность и выделения кластеров среди векторов весов-состояния.

ГОЛЛАНДСКАЯ НОРМАТИВНАЯ БАЗА

Во времена спада была типичной ситуация, когда правительства испытывали большой дефицит бюджета. Присутствие в бюджете не фиксированных заранее статей расходов, например, на социальные выплаты, — с одной стороны, и уменьшение прямых поступлений налогов вследствие снижения темпов экономического развития — с другой, — вынуждало правительства проводить политику финансовых пирамид, которую деликатно называли «расходованием дефицита». Недостачу в бюджете обычно покрывали путем выпуска долговых обязательств, увеличением ставок налогов или тем и другим сразу. При этом всякий раз, когда правительство принимает решение о привлечении капитала, встает вопрос об объемах и сроках.

Традиционно MoF брало ежегодный кредит в 5.7 миллиардов гульденов через центральный банк по льготной ставке. Значение кредита еще более возросло после введения в сентябре 1991 г. новых правил расчета — Primacheque. Действовавший до этого режим Algemene Termijnenwet (ATW) позволял производить платежи не позднее последнего дня месяца включительно при условии, что этот день — рабочий. Если же этот день приходился на выходной, то платежи разрешалось отложить до первого рабочего дня следующего месяца. Чтобы избежать штрафов за просрочку, налогоплательщики

обычно платили вперед, т.е. до последнего рабочего дня. Новое правило Primacheque облегчало для налогоплательщиков работу с наличностью: теперь им гарантировалось, что платежи, сделанные в последний рабочий день месяца, будут оприходованы налоговыми службами до начала первого рабочего дня следующего месяца. Это новшество привело к тому, что, начиная с октября 1991 г., платежи стали совершаться, как правило, в последний день месяца, а не в течение его, как раньше. Такая концентрация платежей резко усилила месячные пики поступления налогов, что, в свою очередь, вызвало увеличение кредитных возможностей центрального банка.

С переходом к единому общеевропейскому рынку подобное «сотрудничество» MoF с центральным банком должно прекратиться. К началу 1994 г. привлечение кредитов должно быть поэтапно сведено к нулю, после чего MoF будет вынуждено действовать на рынке кредитов, подчиняясь законам спроса и предложения. Предвидя возможные неувязки в бюджете, MoF пересматривает имеющиеся у него системы прогноза и ищет новые методы, в частности, связанные с применением нейронных сетей.

Перейдем теперь к обзору традиционных методов оценки налоговых поступлений, а также тех методов, которыми пользуется в своей работе MoF.

ТРАДИЦИОННЫЕ МЕТОДЫ ОЦЕНКИ

Для оценки общего объема поступлений в бюджет широко применяются различные методы, в том числе: эконометрическое моделирование, множественная регрессия, анализ временных рядов.

При эконометрическом подходе оцениваются параметры, входящие в заранее известные уравнения модели экономического поведения. В рамках этого подхода Центральное плановое управление Голландии разработало модель FKSEC [57]. Эта объемная эконометрическая модель предназначена для ежеквартального прогноза ряда общеэкономических показателей. В свою очередь, по этим показателям можно прогнозировать сумму валовых налоговых поступлений. Ежегодно правительство публикует эту оценку (так называемая Миллионен-нота), и она является плановой цифрой для MoF.

Очевидно, что эконометрические модели, будь то FKSEC или модель, предложенная Отеном и Роббом [18], надежны лишь постольку, поскольку надежны и качественны данные о значениях переменных. Подчас в соответствующих рядах имеются пропуски — главным образом, потому, что официальные цифры часто запаздывают или являются предварительными. Пробелы в данных затрудняют прогнозирование, и приходится прибегать к экстраполя-

ции с помощью модели ARIMA, чтобы получить недостающие значения (см. [27]).

Однако, основная трудность здесь состоит в том, что экономические данные (в особенности, в периоды, когда ситуация быстро меняется) содержат гораздо меньше степеней свободы, чем это требуется для оценки параметров модели. Поэтому специалисты, занимающиеся анализом временных рядов, пользуются хорошо специфицированными статистическими моделями со всего одной или двумя переменными. Кроме этого, методы ARIMA и VAR успешно применялись и для непосредственной оценки поступления налогов (см. [24], [202]). Реально MoF Голландии оценивает ежемесячные поступления с помощью модели ARIMA(0,0,0)(0,1,1)₁₂. Для прогнозов задним числом модели временных рядов типа ARIMA-MoF часто оказываются не хуже эконометрических, но у них есть тот недостаток, что эти модели не содержат переменных и соотношений, и, следовательно, по результатам расчетов трудно сделать какие-либо выводы относительно экономической политики.

Фуллerton [117] показал, что хороший результат может получиться при сочетании разных методов. Годовой сбор налогов в штате Айдахо лучше всего прогнозируется с помощью модели, в которой берется взвешенное среднее оценок, полученных по модели ARIMA и по эконометрической модели. Но здесь опять необходимо заранее установить вид уравнений, описывающих связи между переменными, а это непросто в условиях, когда происходят быстрые изменения в экономической обстановке. С учетом всего сказанного, для MoF представляет интерес разработка новых адаптивных методов, которые:

- могли бы работать в условиях, когда ряды значений переменных содержат пропуски;
- не требовали бы априорной спецификации модели;
- могли бы выдавать прогноз быстро меняющихся во времени процессов.

ВЫБОР ПЕРЕМЕННЫХ

Задача состоит в том, чтобы построить модель помесечного прогноза валового сбора налогов (RECEIPT) на период с января 1989 г. по март 1993 г. Брутто-показатель равен нетто-показателю плюс сумма взносов на государственное страхование, которые начисляются на пропорциональной основе из расчета 27% от общей суммы налогов за рассматриваемый период.

Для модели были выбраны 13 переменных, которые описываются ниже. После того, как все они были перемасштабированы так, чтобы

их значения лежали в интервале [0,1], с помощью кросс-корреляционного анализа были исследованы связи целевой переменной (общей суммы налогов) со сдвинутыми назад во времени значениями каждой из переменных.

V1: Календарный эффект (CAL)

Мы уже говорили о том, что пики в конце или в начале месяца объясняются действием закона ATW и правила Primascheque. В модели каждому месяцу приписывается свой код для учета календарного эффекта:

- 0 — если последний день этого месяца — нерабочий и, следовательно, в этом месяце собирается только часть положенных налогов;
- 1 — обычный месяц (без календарных эффектов);
- 2 — месяц, в начале которого происходят поступления от предыдущего месяца, потому что у того последний день был нерабочим, и при этом последний день данного месяца — также нерабочий;
- 3 — месяц, у которого последний день — рабочий, и происходит полный сбор, а в предыдущем месяце — нерабочий.

V2: Официальная оценка валовой годовой брутто-суммы налогов (ANNUAL)

MoF считает официальный правительственный прогноз своим целевым показателем. Помесечную оценку MoF получает так: делит годовую сумму на 12, а затем вносит поправки, отражающие изменения в фискальных правилах и ставках, которые не учитывались в прогнозе правительства. Этот показатель очень важен, и MoF старается уточнить свои помесечные прогнозы.

V3: Сезонность (SEA)

Эта переменная отражает мультипликативную сезонную закономерность, полученную из анализа реальных налоговых поступлений за период с 1978 г. по 1992 г. Здесь используется величина чистых (нетто) поступлений, равная брутто-сумме минус 27% от нее, составляющих взносы в систему государственного страхования. Нетто-показатель используется потому, что именно в этом виде данные публикуются.

V4: Число рабочих дней в месяце (DAY)

Месяцы с большим количеством рабочих дней (23) обычно соответствуют пикам во временных рядах налоговых поступлений, а в месяцы, где число рабочих дней мало (18), сумма налогов бывает

меньше. Кроме того, в «длинные» месяцы больше оборот и, соответственно, комиссионные. Переменная перемасштабируется так, чтобы минимальное значение (18) соответствовало 0, а максимальное (23) — 1.

V5: Совокупное потребление (CON)

Совокупное потребление в частном и общественном секторах влияет на оборот и, следовательно, на суммы акцизных и гербовых сборов. Заметная трехмесячная кросс-корреляция между потреблением и чистым сбором налогов указывает на то, что здесь разумно взять временной лаг в три месяца. Объясняться это может тем, что малые и средние компании перечисляют полученные комиссионные сборы поквартально.

V6: Ставка МБК (AIBOR)

В модель должны быть включены изменения по предполагаемым ежемесячным ставкам предложения на межбанковском рынке Амстердама, поскольку ожидаемая ставка частично определяет потенциальную прибыльность инвестиций (NPV-анализ), а рост ставок может вызвать рост потребления. Мы вычисляли ожидаемое изменение ставки, вычитая ее текущее значение из прогноза на месяц вперед, сделанного по модели ARIMA(0,1,1). Положительное значение означает тенденцию к росту ставок, а отрицательное — к их понижению.

V7: Совокупные вложения в ценные бумаги с фиксированным доходом (INV)

Инвестиции приводят к увеличению комиссионных и поступлений от налогов на доходы по ценным бумагам. Хотя связь между инвестициями и налогом на корпорации с трудом поддается количественному описанию, можно считать, что инвестиции положительно влияют на прибыль и, тем самым, на сумму налогов с корпораций. Наоборот, амортизация может привести к временному снижению доходов, облагаемых налогом. Кросс-корреляционный анализ выявляет большой месячный лаг между инвестициями и чистой суммой собранных налогов (инвестиции берутся со сдвигом один месяц назад).

V8: Уровень безработицы (UNEM)

Эта переменная отражает уровень экономической активности, а также предполагаемые суммы налогов на доходы корпораций и частных лиц.

V9: Амстердамский индекс курсов акций (CBS)

Индекс курсов обыкновенных акций выражает уровень доходов корпораций (и соответствующих налогов). Очевидно, в списке представлена лишь малая часть из всех компаний, действующих в Голландии.

V10: Предложение денег (M1)

Этот узкий показатель денежной массы в обращении влияет на уровень процентных ставок и на потребление. Предполагается, что повышение M1 вызывает увеличение расходов с соответствующим ростом налоговых поступлений.

V11: Показатель фазы цикла деловой активности (CYC)

Валовый национальный продукт с удаленным трендом. Показывает повышения и понижения конъюнктуры.

V12+V13: Погода: осадки + температура (RAIN+TEM)

В модель включены помесечные оценки для выпавших осадков и температуры, поскольку они прямо влияют на потребление энергии и соответствующие комиссионные сборы и косвенно — на склонность потребителей к затратам. Метеорологические факторы вызывают некоторые колебания в налоговых поступлениях, а на абсолютный уровень влияют мало.

НЕЙРОННО-СЕТЕВАЯ МОДЕЛЬ

Эта модель была построена на материале 45 ежемесячных промасштабированных наблюдений за период с января 1989 г. по сентябрь 1992 г. Оставшиеся 6 наблюдений (с октября 1992 г. по март 1993 г.) использовались для проверки модели. К сожалению, MoF не смогло получить более объемную базу данных. В связи с тем, что относительно малый объем данных не позволял использовать подтверждающее множество, результаты, показанные на тестовом множестве,

ве, не являются максимально выверенными в смысле проверки на дополнительных примерах.

В итоге была выбрана модель сети MBPN (13-2-1) с одним скрытым слоем, содержащим два элемента, и прямыми связями между входом и выходом. Коэффициент обучения был взят равным 0.8, а крутизна — равной 1. Далее мы приводим данные по среднеквадратичной ошибке на обучающем и тестовом множествах для различных конфигураций.

Для некоторых конфигураций количество весов явно превосходило число входных данных (наблюдений). Хотя недостаток степеней свободы делает оценку сомнительной, мы приводим здесь результаты работы 13-27-1 модели, чтобы проиллюстрировать доказанную Колмогоровым в 1957 г. и популяризованную Хехт-Нильсеном [137] теорему о существовании отображения. Эта теорема утверждает, что любая непрерывная функция может быть реализована трехслойной нейронной сетью, имеющей во входном слое m (в нашем случае 13) элементов, промасштабированных на $[0,1]$, $(2m+1)$ элементов-процессоров в единственном скрытом слое и n элементов в выходном слое. Таким образом, гарантируется, что иерархическая многослойная нейронная сеть может решить любую нелинейно отделимую задачу и может точно реализовать любое отображение m -мерных входных векторов в n -мерные выходные. При этом теорема ничего не говорит нам ни о возможности реализовать отображение посредством сети меньших размеров, ни о том, что для этого подойдут обычно используемые сигмоидные преобразования.

Тип сети	MSE при обучении	MSE при проверке	Число эпох	Число весов	NBIC
13-2-1	0.021	0.063	45	44	-0.231
13-3-1	0.019	0.066	55	59	0.9639
13-4-1	0.016	0.07	45	74	2.0338
13-5-1	0.019	0.065	55	89	3.4808
13-10-1	0.026	0.063	10	164	10.154
13-27-1	0.021	0.055	22	419	31.477

Таблица 4.1. Сводка результатов для сетей различной конфигурации

Для определения наилучшего размера сети мы пользовались известным правилом для временных рядов, которое называется байесовским информационным критерием (BIC). В случае, когда две модели давали одинаковое качество результатов, предпочтение отдавалось более простой из них, т.е. имеющей меньшее число параметров. Де Гроот и Вуртц [83] предложили модифицированный нормализованный BIC под названием NBIC в виде

$$NBIC = \ln \left(\frac{\sum (\text{Цел.перем.}_k - \text{Прогноз}_k)^2}{N} \right) + \left(\frac{\ln N}{N} \right) p, \quad (1)$$

где $k = 1, 2, \dots, N$, N — число наблюдений в обучающем множестве p — число весов.

Первое слагаемое представляет собой логарифм среднеквадратичной ошибки на обучающем множестве. Второе слагаемое зависит от числа степеней свободы и растет линейно с ростом размера сети. Критерий действует так: сеть, имеющая наименьшее значение NBIC, обладает наилучшими способностями к прогнозу и обобщению. Видно, что из всех испробованных конфигураций наилучшей оказалась сеть 13-2-1.

При относительно простой конфигурации эта сеть имеет 44 потенциальных степени свободы при том, что в обучающем множестве имеются 45 наблюдений. Чтобы смягчить это несоответствие, мы убрали несколько переменных, сильно коррелированных либо с Миллионен-потой (V2), либо с потреблением (V5). Таким образом были исключены следующие переменные:

совокупные вложения в ценные бумаги с фиксированным доходом (V7),
уровень безработицы (V8),
индекс курсов акций CBS (V9),
предложение денег (V10).

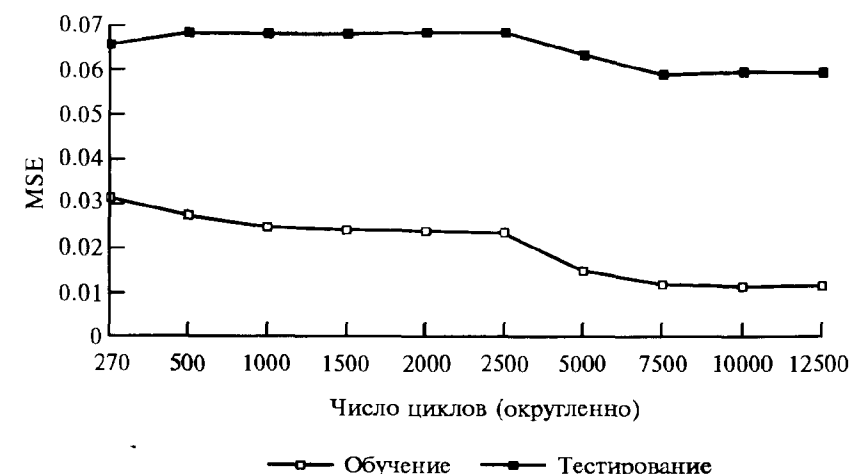


Рис. 4.1. MSE на обучающем и проверочном множествах

В результате получилась архитектура 9-2-1 с 32 параметрами, для которых была проделана повторная оценка. На диаграмме показана

MSE на обучающем и проверочном множествах. Оптимальная длительность обучения составляет примерно 7500 циклов. При дальнейшем увеличении числа эпох в обучении MSE на проверочном множестве начинала медленно расти.

По сравнению с 13-2-1 сетью значения MSE и на обучающем, и на проверочном множествах получаются чуть-чуть лучше. Перед тем, как делать выводы собственно о структуре сети, разумно сравнить ее результаты с такими классическими методами, как многомерная регрессия или модель ARIMA (собственной разработки МоФ).

Наряду с таким хорошим критерием успеха, как MSE, можно пользоваться также так называемой средней относительной дисперсией ARV (см. [275]):

$$ARV = \frac{\sum_{k=1}^N (\text{Цел.перем.}_k - \text{Прогноз}_k)^2}{\sum_{k=1}^N (\text{Цел.перем.}_k - \text{Среднее})^2} = \frac{MSE}{\sigma^2}, \quad (2)$$

где N — число наблюдений, Среднее — среднее значение целевого ряда, σ^2 — дисперсия целевого ряда.

Нормировка MSE устраняет зависимость от динамического разброса данных и учитывает волатильность внутри базы данных. Оценка качества работы сети будет дана в сравнении с результатами регрессии и модели ARIMA. Оба этих метода будут вкратце изложены. Регрессионная модель была построена с использованием только 45 наблюдений из обучающего множества:

$$\begin{aligned} \text{RECEIPT}_t = & \alpha + \beta_0 \text{CAL}_t + \beta_1 \text{ANNUAL}_t + \beta_2 \text{SEA}_t \\ & + \beta_3 \text{DAY}_t + \beta_4 \text{CON}_t + \beta_5 \text{AIBOR}_t + \beta_6 \text{CYC}_t \\ & + \beta_7 \text{RAIN}_t + \beta_8 \text{TEM}_t + \varepsilon_t. \end{aligned} \quad (3)$$

Одна из переменных — ANNUAL, месячная раскладка годового прогноза правительства, имеет здесь особенно большое значение. Это неудивительно, поскольку МоФ рассматривает эту *зависимую* переменную как *цель* и старается приблизить значение целевой переменной (RECEIPT) к правительственному прогнозу (это — пример так называемого условного прогнозирования). Число рабочих дней (DAY), календарные эффекты (CAL) и сезонность (SEA) также играют существенную роль. Погода (RAIN, TEM), ожидаемая процентная ставка (AIBOR) и потребление (CON) существенного влияния на решение не оказывают.

Значение статистического показателя Дарбина–Уотсона указывает на наличие отрицательной корреляции разностей ряда. Поэтому было бы разумно перейти к разностям первого порядка или преобразовать регрессионную модель к такому виду, когда применима процедура OLS (например, итерационный процесс Кокрана–Оркетта). Од-

нако, поскольку МоФ больше заинтересовано в прогнозировании уровня, а не тренда, разности тут плохо подходят. Обычный регрессионный OLS-анализ в применении к прогнозу ежемесячных налоговых сборов дал неудовлетворительные результаты. Поэтому МоФ решило применить одномерный метод Бокса–Дженкинса. Получившаяся в результате модель ARIMA(0,0,0)(0,1,1)₁₂ с параметрами, определенными из того же самого обучающего множества данных, имеет следующий вид:

$$(1 - B_{12}) \ln z_t = 0.04367 + (1 - 0.751 B_{12}) \alpha_t. \quad (4)$$

Здесь B — оператор сдвига назад, а α — составляющая ошибки.

Дисперсионный анализ

	DF	Сумма квадратов	Средний квадрат
Регрессия	9	.85000	.09444
Разность	35	.79126	.02261
F=4.17758		Значимость F=.0010	

Переменная	B	SE B	Beta	T	Sig T
RAIN	-.023820	.110330	-.028191	-.216	.8303
SEA	.152681	.104093	.206986	1.467	.1514
AIBOR	-.064621	.153774	-.058243	-.420	.6769
CON	-.222539	.239240	-.277351	-.930	.3586
CAL	.111763	.079241	.213171	1.410	.1672
TEM	.111333	.100348	.157154	1.109	.2748
CYC	.081578	.092620	.128230	.881	.3844
DAY	.188244	.126605	.246653	1.487	.1460
ANNUAL	.349709	.173124	.575489	2.020	.0511
(КОНСТАНТА)	.060655	.125002		.485	.6305

Значение теста Дарбина–Уотсона = 2.94431

Таблица 4.2. Сводные результаты регрессионного анализа

Сравнительные характеристики всех трех методов видны из сопоставления соответствующих значений ARV (средней относительной дисперсии):

	Регрессия	ARIMA	Сеть
Обучение	0.4821	4.621	0.3165
Проверка	0.8972	3.845	0.7049

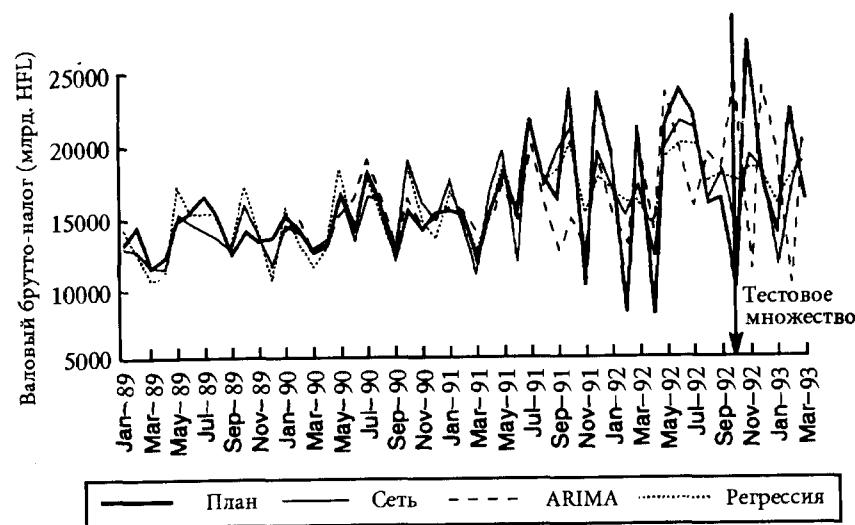


Рис. 4.2. Общая сумма налогов: оценки и действительность

И регрессия, и сеть имеют лучшие характеристики, чем ARIMA. Причина этого в том, что ARIMA является одномерной моделью, где в принципе не могут учитываться календарные эффекты или число рабочих дней. Совокупное действие этих эффектов, начиная с сентября 1991 г., вызывает колебания уровня поступлений налогов от месяца к месяцу и внутри месяцев. Далее, сеть дает более точную оценку, чем регрессия. Причина может быть связана с присутствующей в данных нелинейностью. Значения R -отношения Вигенда¹ 0.705 и 0.743, соответственно, для обучающего и тестового множеств также свидетельствуют о наличии (возможно, слабых) нелинейных связей.

Коль скоро сеть имеет лучшие характеристики, чем модель ARIMA и регрессия, попробуем разобраться в ее внутренней структуре и рассмотрим вклад каждой из девяти фазовых переменных.

ВКЛАД КАЖДОЙ ИЗ ПЕРЕМЕННЫХ ПО ОТДЕЛЬНОСТИ

Структура сети отражается в весах, которые имеют связи, идущие от входных переменных к выходным. Тем самым, величина веса говорит о степени важности данной переменной. Однако до сих пор не разработано никакого систематического способа количественной оценки вклада отдельного веса в выход сети, аналогичного, например, доверительным интервалам в регрессионном анализе. Мы пред-

лагаем анализировать вклады отдельных весов в конечный результат с помощью эвристических методов.

	PE1	PE2	PEOUT
BIAS	-0.76	-0.34	-0.63
CAL	1.26	0.11	0.85
ANNUAL	-1.24	-0.30	0.15
SEA	0.01	-0.31	0.42
DAY	0.76	-0.28	0.73
CON	-1.03	-0.11	-0.76
AIBOR	-0.39	-0.30	-0.29
CYC	-0.27	-0.39	0.08
TEM	0.36	0.04	0.35
RAIN	-0.44	-0.11	-0.30
PE1			-1.56
PE2			0.09

Таблица 4.3. Матрица весов для 9-2-1 сети

Анализ одного отдельно взятого веса не выявляет существа дела, поскольку при этом не учитывается пороговый уровень. Далее, большой вес не обязательно приводит к насыщению сигмоида, так как соответствующие значения входа могут быть очень маленькими. Из матрицы весов 9-2-1 модели с непосредственными связями между входами и выходами видно, что связь между вторым скрытым нейроном (PE2) и выходным элементом (PEOUT) имеет очень малый вес. Однако 9-1-1 сеть уже не способна давать выход того же качества (в смысле коэффициента ARV).

Связи, ведущие от потребления, числа рабочих дней и календарных эффектов к первому скрытому элементу, имеют большие веса (соответственно, -1.03, 0.76 и 1.26). Отрицательные веса непосредственных соединений процентной ставки и дождя с выходным элементом (PEOUT) представляются правдоподобными, тогда как влияние потребления на результат неожиданно происходит с обратным знаком. Правительственный прогноз (ANNUAL) имеет ожидаемый знак в прямой связи с выходом, хотя по результатам регрессии можно было предположить, что абсолютная величина веса будет гораздо больше. Связь со скрытыми элементами имеет отрицательный знак, но, поскольку связь PE1 с PEOUT также отрицательна, итоговое косвенное влияние этой переменной вполне может оказаться положительным.

Рассмотрение отдельных весов вне связи с другими весами ничего не прибавляет к нашему пониманию того, какой в целом вклад вносит в конечный результат та или иная переменная. Различные наборы входных значений будут по-разному преобразовываться систе-

мой весов, поэтому при анализе необходимо учитывать и сами входные значения. Далее мы рассмотрим два эвристических способа такого анализа.

Прежде всего мы измеряем чувствительность решения к изменениям значения одной входной переменной. Здесь пока каждая переменная учитывается изолированно. Для того чтобы учесть их возможное взаимодействие, мы во втором методе, независимо друг от друга, оцениваем уровни активации всех скрытых элементов. Наблюдения, принадлежащие одному кластеру, дают сходные результаты. При подходе, основанном на оценке чувствительности, измеряется вклад переменных в величину остатка для каждого наблюдения. Вся процедура происходит так:

- 1 Все значения одной входной переменной x заменяются на ее безусловное математическое ожидание (т.е. среднее значение).
- 2 Выход сети переисчисляется для сокращенной матрицы входов и при той же матрице весов.
- 3 Абсолютный остаток полученного выхода сравнивается с абсолютным остатком выхода, полученного на исходной входной матрице, а именно:

$$\text{Вклад}_{\text{var } x} = \left| \frac{(\text{Цел. перем.} - \text{Прогноз}_{\text{нов}})}{(\text{Цел. перем.} - \text{Прогноз}_{\text{исходн.}})} \right|.$$

Если величина этого отношения меньше единицы, то это значит, что исходный остаток больше, чем повторно вычисленный, и переменная влияет на результат с отрицательным знаком. Наоборот, если дробь больше единицы, то эта переменная способствует уменьшению остатка, и ее вклад в искомое решение положителен. Значение дроби около 1 (например, в интервале [0.75, 1.25]) соответствует нейтральному вкладу.

- 4 Прodelать все перечисленные выше действия для всех других переменных.

В табл. 4.4 представлена статистика по видам влияния переменных на результат, полученная на материале обучающего множества. Из нее видно, что наибольшее влияние на прогноз оказывают годовой прогноз правительства, число рабочих дней и календарные эффекты.

	CAL	ANNUAL	SEA	DAY	CON	AIBOR	CYC	TEM	RAIN
Положительное	17	29	16	19	15	7	15	14	11
Нейтральное	16	5	20	13	22	33	20	19	32
Отрицательное	12	11	9	13	8	5	10	12	2

Таблица 4.4. Влияние отдельных переменных (по видам)

Хотя в целом переменные RAIN и AIBOR ведут себя индифферентно, обращают на себя внимание случаи, когда их вклад положителен (таких, соответственно, 11 и 7). Эти случаи заслуживают дальнейшего исследования. С другой стороны, годовой прогноз, который по результатам регрессии играет значительную роль, в 11 случаях неожиданно дает отрицательный вклад.

Можно сделать вывод о том, что здесь желателен дальнейший анализ, который включал бы в себя выявление различных типов входных конфигураций и формирование на их базе архетипов. Разработка такой типологии может привести к выводу правил, с помощью которых MoF сможет более точно предсказывать уровень валовых налоговых поступлений.

Подход, основанный на исследовании архетипов, соответствующих кластерам похожих наблюдений, обладает тем достоинством, что он является многомерным, и в этом состоит его отличие от анализа весов и влияний. Горман и Сейновски [127] предложили способ кластеризации наблюдений с помощью весовой матрицы, зависящей от весов соединений, идущих от входных элементов к исследуемому i -му скрытому элементу. Для k -го наблюдения метрический вектор, или вектор весов-состояния, Q имеет вид $Q_j^{(k)}[i] = [w_{ij} p_j^{(k)}]$, где p — выход j -го входного элемента. Затем для каждой пары векторов весов-состояния вычисляется евклидово расстояние между ними, и все они записываются в матрицу расстояний. На последнем этапе к этой матрице применяется метод иерархической кластеризации. Наблюдения с близкими векторами весов-состояния образуют кластер. Усредняя все наблюдения, принадлежащие одному кластеру, получаем центроид этого кластера. Все центроиды могут быть упорядочены по уровню выходного сигнала или по уровню активации нужного скрытого элемента. Вся процедура продлевается независимо для каждого скрытого элемента. В нашем случае конфигурация сети имеет два скрытых элемента, так что описанный шаг повторяется дважды. Особенно важны значения тех центроидов, которые наиболее сильно активируют данный элемент. Поскольку вес связи, идущей от второго скрытого элемента (PE2) к выходному (PEOUT), очень мал (0.09), с учетом того факта, что элемент PE1 не всегда насыщается (уровень активации меняется в интервале [0.01, 0.77]), мы при дальнейшем анализе элемент PE2 из рассмотрения исключим.

Подход, основанный на рассмотрении векторов весов-состояний, хорошо подходит для сетей такой архитектуры, где имеется только один скрытый слой и нет непосредственных связей между входом и выходом. Поскольку конфигурация нашей сети не удовлетворяет второму требованию, подход был несколько видоизменен.

CENTROID	CAL	ANNUAL	SEA	DAY	CON	AIBOR	CYC	TEM	RAIN	ACTIVATION
C6(4)*	0.00	0.64	0.41	0.55	0.72	0.40	0.61	0.60	0.48	0.01
C5(9)	0.43	0.73	0.56	0.67	0.74	0.37	0.61	0.49	0.33	0.04
C8(1)	1.00	0.78	0.00	0.80	1.00	0.42	0.68	0.39	0.49	0.07
C3(6)	0.00	0.09	0.37	0.47	0.36	0.42	0.42	0.58	0.40	0.10
C7(7)	0.92	0.68	0.52	0.57	0.55	0.37	0.68	0.68	0.28	0.18
C4(1)	0.73	0.17	0.13	0.00	0.04	0.85	0.84	0.49	0.31	0.22
C2(10)	0.43	0.07	0.53	0.66	0.35	0.24	0.37	0.58	0.39	0.35
C1(7)	1.00	0.10	0.68	0.80	0.26	0.25	0.50	0.62	0.23	0.77

*в скобках указано число наблюдений

Таблица 4.5. Значения центроидов (промасштабированные) и соответствующие значения активации (упорядочен)

По результатам кластерного анализа в обучающем множестве было выделено 8 кластеров. Были вычислены центроиды, затем они были упорядочены по возрастанию величины активации. Две пары кластеров (6&5 и 2&1), составляющих вместе 30 наблюдений, относятся, соответственно, к очень низким и очень высоким выходным значениям первого скрытого элемента. Далее, наблюдения выявили для кластеров групп 5 и 6 высокие, а для кластеров 1 и 2 — низкие реальные значения целевой переменной.

Нетрудно определить причинные факторы, вызывающие такие значения активации. Для каждой пары кластеров мы нарисовали (в промасштабированном виде) значения девяти входных переменных. В обоих парах значения переменных для обоих центроидов давали примерно одну и ту же картину.

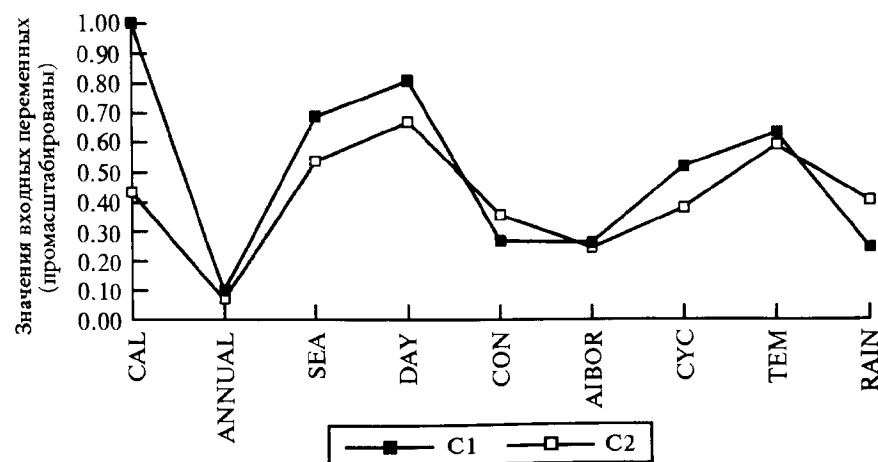


Рис. 4.3. Значения центроидов для кластеров 1 и 2, дающих большие значения активации скрытого элемента PE1

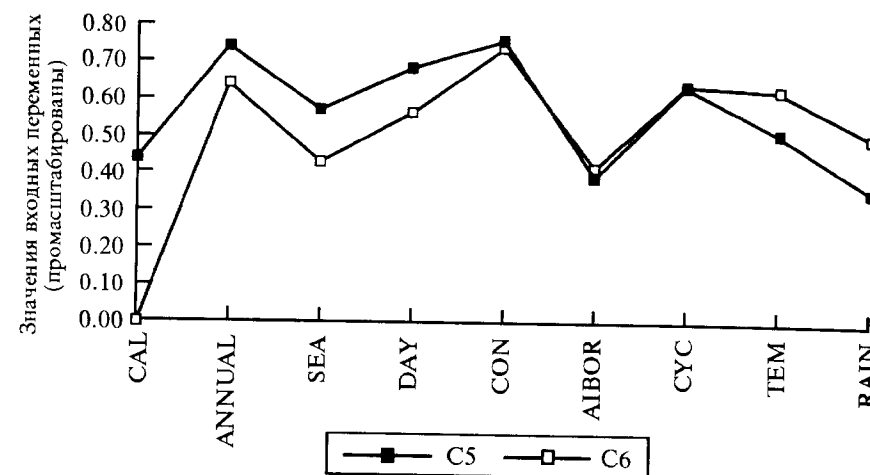


Рис. 4.4. Значения центроидов для кластеров 5 и 6, дающих низкие значения активации скрытого элемента PE1

Низкие значения годового прогноза и потребления подавляют выходной сигнал скрытого элемента. В то же время, большие значения календарного эффекта, числа рабочих дней и температуры стимулируют выходной сигнал элемента PE1, идущий к элементу PEOUT. Заметьте, что группы наблюдений, образующие кластеры 5 и 6, соответствуют низкому целевому показателю.

Помимо сигналов, приходящих от элементов PE1 и PE2, выходной элемент принимает также 9 сигналов, идущих непосредственно от входных элементов. Как неявно заметили Горман и Сейновски, эти сигналы тоже могут быть разбиты на кластеры и упорядочены по уровню активации выходного элемента. В результате такого анализа количество кластеров может оказаться другим, причем в один кластер, вообще говоря, будут входить элементы разных групп. Поэтому определить суммарный эффект, который оказывают выходные сигналы скрытых и входных элементов, будет довольно трудно.

Чтобы упростить задачу, мы для каждого из восьми кластеров вычислили усредненный входной вектор, и вместо исходной базы данных взяли эти усредненные векторы. Затем были получены и рассортированы соответствующие уровни активации выходного элемента. Теперь усредненные входные векторы кластеров 5 и 6 порождают очень высокие значения активации элемента PEOUT (0.993 и 0.996), а усредненные векторы кластеров 1 и 2 дают наименьшие значения активации (0.48 и 0.77). По-видимому, такое изменение порядка говорит о том, что прямые связи (квазилинейные функции) систематически порождают завышенную оценку целевой переменной

(RECEIPTS), а соединения, идущие через узел PE1, уменьшают этот порог.

Теперь, соединяя вместе анализ матрицы весов, учет положительных и отрицательных влияний и выходных значений кластеров, мы в состоянии оценить значимость вклада каждой из переменных. В случае, если разные подходы приведут к разным выводам, следует, скорее, доверять результатам кластерного анализа в силу его многомерной природы.

Календарные эффекты (V1) и число рабочих дней (V4) сильно влияют на выход сети, что подтверждается большими значениями весов их прямых соединений с выходом (0.85 и 0.73), а также кластерным анализом их соединений, проходящих через PE1 (центроиды 1 и 2). Поскольку эти кластеры объединяют наблюдения с низким действительным поступлением налогов, можно считать, что эти две переменные вызывают уменьшение выхода. На интуитивном уровне это понятно: возможности отсрочки платежей отрицательно сказываются на поступлении налогов. Аналогично, опосредованное влияние температуры (V12) на целевую переменную отрицательно, хотя ее суммарный вклад в выходной сигнал остается неясным. Годовой прогноз правительства (V2) влияет на выход резко положительно — все варианты анализа показывают, что высокие значения целевой переменной приходится на высокие значения переменной V2. Но, поскольку точного совпадения в значениях этих двух переменных нет, MoF не может с абсолютной уверенностью положиться на правительственный прогноз при управлении своими активами. Здесь также играет роль то, каким образом MoF раскладывает годовой прогноз по месяцам. Далее, кластерный анализ выявляет положительную связь между потреблением (V5) и выходным значением. Большой отрицательный (−0.76) вес прямой связи невозможно объяснить экономическими соображениями, а общий результат двух влияний может указывать на нейтральность действия этой переменной. Сезонность, по-видимому, влияет на выход линейно (вес = 0.42) с небольшими побочными эффектами. Ее общий вклад примерно в 50% случаев положителен. Но эта переменная не может быть однозначно связана только с большими либо только с малыми выходными значениями. Цикличность (V11) имеет слабое прямое влияние (вес 0.08) и слабое побочное действие. Роль остальных переменных — AIBOR (V6) и дождя (V13) — с трудом поддается интерпретации. Их окончательное влияние нейтрально, и остается неясным, каким образом опосредованные связи компенсируют довольно большие отрицательные веса прямых связей (−0.29 и −0.3).

ВЫВОДЫ

Нейронная модель показывает лучшие результаты, чем ARIMA и OLS-регрессия как на обучающих, так и на новых данных. При этом, однако, из-за небольшого объема базы данных не удастся применить полноценное подтверждающее множество, и поэтому обучающее множество было оптимизировано по отношению к тестовому множеству.

Была сделана попытка выяснить структуру реализуемого отображения. Исследование вклада каждой переменной через веса непосредственных связей и через значения векторов весов-состояния позволило получить представление как о линейных, так и о нелинейных компонентах модели. По-видимому, календарные эффекты, число рабочих дней и температура влияют на значения целевой переменной отрицательно, тогда как увеличение годового правительственного прогноза и потребления приводит к росту целевого значения. Остальные переменные активны при любых значениях целевой переменной. Такое сложное влияние переменных едва ли уловимо средствами регрессионного анализа. Различие между линейными и нелинейными компонентами, возможно, несколько преувеличено, так как значение отношения R больше 0.7.

В настоящее время Министерство финансов пробует различные способы улучшения нейронно-сетевой модели. Показатели работы сети могут улучшиться после разбиения целевой переменной на составные части, потому что, например, на размер налога с продаж влияют одни факторы, а на размер налогов на наследство — совсем другие. Другой способ состоит в том, чтобы свести целевую переменную к совокупности банковских счетов. Так как разные группы налогоплательщиков перечисляют налоги на разные счета, каждый из этих счетов имеет свои особенности, которые можно учесть. Далее, можно включить сюда сдвинутые назад значения целевой переменной или разности. Еще одна возможность состоит в том, чтобы включить в число входов оценку, полученную с помощью модели ARIMA. Наконец, следуя идее Фуллертон, из прогнозов, полученных разными методами, можно тем или иным способом составить один более надежный.

ПРИМЕЧАНИЯ

¹ R = дисперсия остатков нелинейной модели/дисперсия остатков линейной модели.

Временные ряды в задачах расчета цен опционов европейского типа

Новые методы, в том числе методы нейронных сетей, дают возможность исследовать нелинейные модели, ранее не подвергавшиеся тестированию. Возможно, что традиционные модели формирования цен оказываются недостаточно хорошими именно из-за неадекватной спецификации, а не из-за свойств эффективности рынка. В этой книге мы исследуем вопрос о том, можно ли с помощью MBPN-модели получить возможности для извлечения прибыли на небольшом отрезке времени. Используя базу данных о сделках, совершаемых в течение рабочего дня на Европейской бирже опционов в Амстердаме, мы пытались прогнозировать размер прибыли по обыкновенным акциям компании Филипс. Две нейронные сети и обычный линейный регрессионный анализ сравнивались между собой по трем критериям: средней квадратичной ошибке (MSE), ρ и полученному доходу. По всем критериям лучшие результаты показала адаптивно обученная 33-14-1 сеть, которая по простой однопериодной торговой стратегии дала доход 11% в пересчете на годовые.

ПОСТАНОВКА ЗАДАЧИ

Использование нейронных сетей для изучения нелинейных моделей формирования цен акций вносит ясность в вопрос о том, в какой степени недостатки линейных моделей вызваны их неадекватной спецификацией, а в какой — предположением об эффективности рынка.

Во многих публикациях последних лет исследуется наличие структуры «следования с запаздыванием» в ценах на опционы и обыкновенные акции. В большинстве этих работ делается вывод о том, что цена опционов является ведущим ориентиром, которому следует цена соответствующей бумаги на наличном рынке. Из этого следует такой очевидный вывод в отношении гипотезы эффективного рынка (она обсуждалась в гл. 3): зная положение на рынке опционов, можно извлечь информацию, которая еще не дошла до рынка наличности. Удачный прогноз будущей цены дает возможность для проведения выгодной торговой стратегии. В этой главе мы применим нейронные сети для прогноза будущего дохода по акциям транснациональной электронной компании Филипс, котирующимся на Амстер-

дамской фондовой бирже. В качестве входов для нашей сети мы будем использовать различную информацию о позициях, занятых по апрельским 1992 г. опционам колл на акции Филипс на Европейской бирже опционов (ЕОЕ).

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ

Возможность точного предсказания ситуации на рынке представляется проблематичной, и разработанные для этого модели дают неудовлетворительные результаты. Очевидно, что такие модели не могут описывать действительно эффективный рынок, где вся поступающая информация мгновенно сказывается на ценах. Если предположить, что гипотеза эффективного рынка верна и изменение цены акций представляет собой случайное блуждание, то ни фундаментальный, ни технический анализ не имеют под собой никаких оснований. Любая предсказуемая возможность для извлечения прибыли будет использована задолго до того, как аналитик проделает свои вычисления. Почему же так много отдельных интеллектуальных индивидуумов и инвестиционных компаний продолжают заниматься прогнозами и заключают сделки «против рынка»? Почему солидные банки тратят столько усилий на составление и публикацию ежемесячных и еженедельных прогнозов состояния экономики и финансов, если те же самые, а то и лучшие результаты можно получить с помощью датчика случайных чисел? Почему управляющие портфелями ценных бумаг столько работают над отбором акций для своего портфеля при том, что эти же люди как физические лица не стали бы играть на превышение индекса (т.е. формировать инвестиционный портфель, растущий в цене быстрее, чем в среднем весь рынок)?

Частичный ответ на эти вопросы дает традиционная экономическая теория финансов, основывающаяся на модели ценообразования на рынке капиталовложений (CAPM) и таких ее обобщениях, как теория арбитражного ценообразования (АРТ), см. об этом далее в гл. 6. Систематический риск, связанный с определенной позицией на фондовом рынке, должен иметь компенсацию. Безрисковый доход для всех одинаков, а дальше теория утверждает, что чем выше риск, тем больше должен быть доход от инвестиций. Совместный анализ поведения цены акций и экономических факторов позволяет выявить закономерности в движениях цен и показателей риска. Таким способом может быть предсказана часть дохода — так называемая премия за риск. Если премия за риск превысит или, наоборот, упадет ниже ожидаемого среднего значения для этого типа риска, то возникает некоторый аномальный доход, что указывает на появление новой существенной информации.

Примитивная гипотеза эффективного рынка может быть с очевидностью опровергнута, если рынок «зашумлен» наличием неравновесных цен, психологическими барьерами и возвратными движениями цен при достижении крайних значений (см. [279]). Где та грань, за которой способность эксперта выбрать прибыльную акцию перестает быть просто удачей и становится системой? Такие новые технологии, как нейронные сети, позволяют получать возможности для извлечения прибыли на ограниченных отрезках времени и/или на краткосрочных контрактах. Даже если предположить, что рынок рационален, т.е. что все индивидуальные инвесторы ведут себя в среднем рационально, это вовсе не значит, что не может существовать не замеченных ранее закономерностей в ценах и различных индикаторах рынка.

Поскольку новые методы позволяют исследовать ранее не использовавшиеся модели с нелинейной спецификацией, возможно, что преимущества этих методов перед традиционными объясняются как раз более точной спецификацией, а не свойствами эффективности рынка.

Возможно, самый известный и признанный теоретический результат в области финансов — это модель Блэка–Шоулса для определения цены опционов (OPM, Option Pricing Model). Согласно этой модели, цена опциона прямо определяется предсказуемыми показателями наличного рынка соответствующих основных ценных бумаг. Поэтому очень интересно было бы выяснить, существуют ли связи типа запаздывания между ценами опционов и ценами на наличном рынке. Характерной особенностью опционов является то, что небольшие начальные вложения позволяют получать прибыль от изменения рыночных курсов, соответствующую большому количеству акций (так называемый левередж). Пантон [210] высказал мысль, что эти соображения «ликвидности» должны приводить к тому, что цены акций будут следовать за ценами опционов. За большинством сделок по опционам рано или поздно следуют сделки по соответствующим акциям — в частности, потому, что издатели (продавцы) опционов немедленно хеджируют свои позиции сделками на рынке акций (дельта-хеджирование), а также потому, что многие контракты исполняются раньше срока (там, где в ходу опционы американского типа). В результате та информация, на основании которой принимаются решения по сделкам с опционами, в некотором преобразованном виде передается на рынок акций.

Опционы дают инвестору наилучшие возможности для управления акционерным риском. Если кто-то из участников рынка предвидит его возможное потрясение (зависящее от наступления определенных событий), то это отразится на структуре позиций рынка опционов, а на рынке наличности может никак не проявиться. Блэк

[42] показал, что в определенных случаях опционный эквивалент позиции по акции может иметь для инвестора иную (более высокую или более низкую) неявную ценность, чем сама акция. Это различие может объясняться всевозможными издержками на совершение сделок (так называемыми транзакционными издержками), ограниченным капиталом для совершения наличных сделок, преимуществами в отношении дивидендов и налогов. Таким образом, причины того, что инвестор, располагающий новой информацией, скорее предпочтет воспользоваться ею на рынке опционов, чем на рынке акций, лежат в институциональном устройстве рынка. В выполненном по заказу Амстердамской фондовой биржи исследовании положения дел в Голландии Фэйз и др. [109] приводят три причины, почему рынок опционов может быть более привлекателен, чем рынок акций:

- меньшие требования к марже и более низкие транзакционные издержки,
- более непосредственный контакт спроса и предложения (торговля с голоса вместо экранной системы котировок),
- более быстрое администрирование.

О том, что цены акций следуют за ценами опционов, говорят результаты нескольких исследований. Манастер и Рендлман [185] обнаружили этот эффект на материале ежедневных данных о торгах, сравнив доход «от закрытия до закрытия» по портфелям опционов, основанных на различии действительных и предполагаемых (по модели Блэка–Шоулса) цен на акции. Авторам удалось установить, что цены закрытия на опционы несут в себе наиболее свежую информацию, еще не учтенную в ценах на акции. Многие рынки опционов, и в том числе ЕОЕ, закрываются позже, чем фондовые биржи (ЕОЕ — на 10 минут). Эта важная подробность, конечно же, сказывается на причинно-следственных связях между ценами закрытия на обоих рынках. Впрочем, в рассмотренной ниже нейронно-сетевой модели мы имеем дело с данными за один торговый день, а на них эта сторона дела сказывается в меньшей степени.

Бхаттачария [41] исследовал связи между ценами на опционы и на акции не только день ото дня, но и в течение одного биржевого дня. Он описывает модель торговой стратегии, показывающую, что за счет внутридневных позиций нельзя получить дополнительную прибыль. Напротив, позиции, переходящие со дня на день, могут принести прибыль, что было продемонстрировано моделированием овернайт-стратегий. Спецификация модели Бхаттачария предполагает, что рынок опционов влияет на рынок акций, а обратного влияния или одновременного их взаимодействия нет. При этом его результаты согласуются с тем, что получили Манастер и Рендлман.

Энтони [14] обнаружил связи между ценами на опционы колл и на акции и сделал вывод (согласующийся с выводами Манастера-Рендлмана и Бхаттачария) о том, что торговля опционами опережает торговлю акциями на один день. Однако более тщательное исследование результатов показывает, что этот временной лаг присутствует не всегда, и в некоторых случаях рынок акций даже может опережать рынок опционов.

Этот эффект обнаружили Стефан и Уэли [249], когда они с помощью множественной регрессии сравнивали внутрисдневные цены на акции, наведенные ценами на опционы, с действительными ценами акций. При этом обнаружилась обратная зависимость с лагом 15 минут. Авторы предлагают объяснение этого явления, основанное на специфических закономерностях, характерных для торгов в течение одного дня. Обычно сразу после открытия биржи интенсивность торгов бывает большая, и к полудню она постепенно спадает. После обеда объем значительно растет и затем резко падает незадолго до закрытия. Характер торгов по опционам иной. При открытии торгов по опционам, которое происходит спустя 45 минут после открытия торгов по акциям, наблюдается резкий рост объема торговли. Затем объем в течение всего дня уменьшается, следуя за объемом торговли акциями. После закрытия фондовой биржи интенсивность торгов по опционам возвращается к высокому уровню, имевшему место при открытии. Ясно, что во время двух пиков — при открытии и при закрытии биржи опционов — держатели опционных позиций уточняют их, основываясь при этом на наиболее свежей информации, содержащейся в ценах акций.

Мы можем сделать вывод, что рынок опционов и рынок соответствующих акций, несомненно, связаны друг с другом, хотя по вопросу о направленности и характере этой связи есть различные мнения. Поэтому для нейронной сети, рассчитанной на прогнозирование поведения цен акций, данные о позициях на рынке опционов представляют собой важный источник потенциальной информации.

В следующем разделе будут описаны эксперименты с MBPN-сетями, в которых на основании данных о сделках по апрельским 1992 г. опционам колл на акции Филипс, совершаемых в течение дня на ЕОЕ, дается прогноз дохода по этим акциям на 15 минут вперед. Крупная многонациональная электронная компания Филипс была выбрана в том числе и потому, что эта фирма в течение первого полугодия каждого года не выплачивает никаких дивидендов. Благодаря этому мы могли использовать ОРМ-модель Блэка-Шоулса без внесения поправок на дивиденды. Кроме этого, поскольку акции Филипс высоколиквидны как на наличном рынке, так и на рынке производных финансовых инструментов, база данных включает в себя достаточно большое количество сделок по опционам всех серий.

ЭНДОГЕННЫЕ И ЭКЗОГЕННЫЕ ПЕРЕМЕННЫЕ

Для прогнозирования значений единственной выходной переменной — дохода по акциям через 15 минут — мы использовали 33 экзогенные переменные. Поскольку использовались данные о реальных сделках, цены в исходном файле соответствуют нерегулярно расположенным моментам времени. Поэтому мы брали усредненные значения цены за каждый 15-минутный интервал. Так как ЕОЕ открывается в 9.30 утра, а закрывается в 5 ч. вечера, каждый день торгов состоит из 30 таких интервалов. Для того чтобы уменьшить влияние периодов вялой торговли, мы исключим из дальнейшего анализа данные, относящиеся к промежутку времени с 4.45 до 5 часов. Поправок на обеденный перерыв не делалось. В случае, если в течение 15-минутного интервала никаких сделок по опционам не происходило, предполагалось, что цена акций на наличном рынке в течение такого времени останется постоянной и равной простому (невзвешенному) среднему значению, зафиксированному на предыдущем интервале.

Каждое наблюдение представляет собой совокупность четырех записей (строк в таблице) — цен апрельских 1992 г. опционов колл с четырьмя разными ценами исполнения. Целевая переменная есть средний доход по акциям Филипс, полученный в течение последующих 15 минут торгов. Поскольку цена акции (и доход по ней) не зависит от цены исполнения, целевая переменная имеет на каждом интервале одно значение для опционов всех четырех серий.

Все входные переменные можно разделить на две группы: переменные, *распознающие состояние*, которые часто, но не всегда, принимают дискретные значения, и *описательные* переменные, принимающие непрерывные значения. По переменным состояние сеть может распознавать опционы разных серий. В зависимости от значений этих переменных сеть должна по-разному «направлять» входной сигнал. Например, от того, состоялась ли некоторая сделка по опционам в 4 ч. дня в пятницу или в 11 ч. утра в среду, зависят те выводы, которые можно из этого сделать в отношении цены акций. Ситуация, когда модель ведет себя по-разному в зависимости от значений одной или нескольких переменных, — это типичный пример нелинейных взаимодействий. То же самое может относиться и к недискретным переменным. Например, сделка по опциону, до истечения срока которого остается более трех месяцев, и по опциону, до исполнения которого остается совсем немного времени, по-разному влияют на изменения цены акций в краткосрочной перспективе. Если мы смотрим на входные переменные таким образом, то мы неявно подразумеваем, что существует некоторая совокупность решающих пра-

вил, по которой логическим путем определяется текущее состояние системы, с соответствующими выводами относительно значений целевой переменной (или переменных).

Помимо переменных, распознающих состояние, среди входов имеются описательные переменные, которые могут быть связаны с целевой переменной той или иной аналоговой (непрерывной) связью: линейной, квадратичной или обратной. Понятно, что могут существовать переменные, которым присущи оба свойства, например, есть основания считать, что пока процентная ставка находится ниже определенного уровня, она не влияет на цену акции, а выше этого уровня — прямо определяет ее. Следуя этим принципам, мы выбрали входные переменные модели следующим образом:

Цена исполнения (EXERP)

Цена исполнения апрельских 1992 г. опционов колл на акции Филлипс введена в модель для того, чтобы различать опционы разных серий (описанная ниже переменная TTM имеет то же назначение). Эта переменная позволяет узнать, является ли опцион «в деньгах» или «без денег» («in-» или «out-of-the-money») и (в совокупности с другими переменными) определить временную ценность опциона. Из шести имеющихся в обороте серий опционов мы выбрали четыре, объем сделок по которым наибольший. Соответствующие цены исполнения равны 30.00, 32.50, 35.00 и 37.50 флоринов.

Цена опциона (WAVGPOP)

В упомянутой выше модели Блэка–Шоулса (B&S) ОРМ эта переменная нелинейно зависит от пяти других входных переменных: цены акции (S), волатильности (σ), процентной ставки (r), времени до исполнения и цены исполнения. Отклонение реальной цены опциона от теоретической B&S-цены (которая широко используется как эталон для определения цены опциона), может нести в себе определенную дополнительную информацию. Эта переменная играет примерно ту же роль, что и рассмотренная ниже подразумеваемая волатильность (IMVOLEUR). Вероятная связь здесь мыслится такой: если цена опциона колл высокая, то цена акции должна расти. В качестве значений переменной брались средние значения цены апрельских 1992 г. опционов колл за очередные 15 минут. При усреднении учитывался объем каждой сделки (число проданных/купленных контрактов).

№	Обозначение	Описание
1	EXERP	Цена исполнения апрельского 1992 г. опциона колл (4 серии)
2	WAVGPOP	Последовательность усредненных цен апрельских 1992 г. опционов колл
3	TIME	Временная ценность апрельского 1992 г. опциона колл для каждой из цен исполнения
4	TRADAY	День недели, в который происходит сделка
5	TRAHOUR	Час, в который совершается сделка
6	OICA	Открытый интерес по опционам колл
7	OIPU	Открытый интерес по опционам пут
8	CACONAP	Число сделок по апрельским 1992 г. опционам колл, заключенных за 15 минут
9	CACONJU	Число сделок по июльским 1992 г. опционам колл, заключенных за 15 минут
10	CACONOC	Число сделок по октябрьским 1992 г. опционам колл, заключенных за 15 минут
11	ASBI	Разница между ценами предложения и спроса (бид-аск спрэд) апрельских 1992 г. опционов колл
12	RETLAG	Доход по акциям за предыдущие 15 минут
13	TTM	Число дней до исполнения опционов, подсчитанное по Бодде [45]
14	IMVOLEUR	Подразумеваемая волатильность апрельских 1992 г. опционов колл
15	IMVOMADI	Разность между подразумеваемой волатильностью и ее однодневным скользящим средним
16	HISVOLA	Историческая волатильность курса акций на очередном 15-минутном интервале
17	EURO	Процентная ставка Евро для каждого из сроков погашения опционов, в пересчете на годовые
18	IMPLRE	Подразумеваемая годовая процентная ставка по Нуману [206]
19	MMPOSLO	Соотношение длинных и коротких позиций на рынке соответствующих акций
20	THEPOP	Теоретическая цена апрельских 1992 г. опционов колл
21	IQ	IQ апрельского 1992 г. опциона колл
22	CAPUDIF	Разность между количествами заключенных сделок по опционам колл и по опционам пут всех серий
23	TRANS	Число сделок за 15-минутный промежуток времени по опционам, по каждой серии
24	IOCADIF	Разность между количествами in- и out-of-the-money контрактов типа колл
25	IOPUDIF	Разность между количествами in- и out-of-the-money контрактов типа пут
26	DELTAEUR	Коэффициент хеджа для апрельских 1992 г. опционов колл для каждой из цен исполнения
27	OPELASEUR	Эластичность апрельских 1992 г. опционов колл, для каждой из цен исполнения
28	MONEY	Мера того, насколько апрельские 1992 г. опционы колл являются in-the-money
29	GAMMAEUR	Коэффициент гамма для апрельских 1992 г. опционов колл, для каждой из цен исполнения
30	LAMBDAEUR	Коэффициент лямбда для апрельских 1992 г. опционов колл, для каждой из цен исполнения

№	Обозначение	Описание
31	RHOEUR	Коэффициент rho для апрельских 1992 г. опционов колл, для каждой из цен исполнения
32	THETAEUR	Коэффициент theta для апрельских 1992 г. опционов колл, для каждой из цен исполнения
33	DELQU	Число раз, когда менялась котировка апрельских 1992 г. опционов колл с этой ценой исполнения

Таблица 5.1. Входные переменные модели цены акций

Временная ценность опциона (TIME)

Среднюю цену опциона колл (WAVGPOP) можно рассматривать как сумму его внутренней ценности и временной ценности. Последняя включает в себя элемент ожидания. TIME вычислялась отдельно для каждой цены исполнения путем вычитания внутренней ценности опциона (которая равна $S - X$) из его цены c .

День недели (TRADAY)

Эта дискретная переменная используется в качестве входной потому, что иногда наблюдается «эффект дня недели», когда позиция, занятая по опциону, оказывает различное влияние на последующее поведение цены акций в зависимости от того, в какой день недели совершена сделка.

Включая день недели во входной сигнал, мы явным образом учитываем тот факт, что позиция, занятая по опциону в один день недели, может сказываться на цене акций иначе, чем та же позиция, занятая в другой день. Выбор дня может быть связан с раскрытием новой информации. Например, центральные банки предпочитают делать важные шаги в те моменты, когда позиции не могут меняться. Как правило, это — выходные или непосредственно следующие за ними дни. Вследствие этого подразумеваемая волатильность (еще одна из входных переменных) бывает относительно высокой по понедельникам: в этот день рынок усваивает много новой информации. При этом, однако, рынок соответствующих акций ведет себя совсем не так, как это бывает при такой же волатильности в другие дни, — это типичный пример нелинейной связи. Использовались следующие обозначения: понедельник — 2, вторник — 3 и т.д.

Час совершения сделки (TRANOUR)

Смысл здесь тот же, что и с переменной DAY. Так, например, в течение ночи биржа закрыта и новая информация не может влиять на цены до тех пор, пока торги не возобновятся утром следующего дня.

Далее, характер влияния других переменных зависит от времени дня. Например, при открытии торгов на фондовом рынке обычно наблюдается высокая их интенсивность, а затем, в течение утра, она спадает. После обеда объем сделок сильно возрастает, а к закрытию — резко падает. В такой ситуации переменная TRANOUR может оказаться полезной для интерпретации данных об объеме сделок. Сделки, совершенные до 10 ч. утра, обозначаются цифрой 9, с 10 до 11 ч. — 10 и т.д.

Открытый интерес по опционам колл и пут (OICA, OIPU)

Эта переменная выражает общее число позиций по апрельским 1992 г. опционам (отдельно по сериям), оставшихся открытыми на конец предыдущего дня торгов. К сожалению, данные о том, как меняется этот показатель в течение дня, не публикуются, но и ежедневные данные представляют определенную ценность для прогнозирования. Большие значения этого показателя могут указывать на высокую ликвидность рынка опционов и на уверенность в хороших перспективах для соответствующих акций. Далее, участники рынка, располагающие внутренней информацией, предпочитают работать с опционами тех серий, по которым OI наибольший. Низкое значение OI само по себе не несет большой информации о состоянии торговли по опционам данной серии, однако эта переменная может быть полезна в комбинации с количеством сделок (TRANS) и объемом контрактов (CACONAP).

Объем сделок (CACONAP, CACONJU и CACONOC)

Эти три переменные (AP соответствует исполнению в апреле, JU — в июле, OC — в октябре) показывают число сделок по опционам колл, заключенных в течение каждого 15-минутного интервала. Оживленная торговля, скажем, октябрьскими контрактами может привести к тому, что так называемые споттеры рынка опционов станут совершать сделки на наличном рынке, влияя тем самым на будущий доход по акциям.

Доход по акциям (RETLAG)

Средний доход по акциям за предыдущий 15-минутный отрезок времени является исходной точкой для прогнозирования будущего дохода по акциям. Если доход по акциям (и, следовательно, их цена) обладают какой-либо ценностью для прогноза будущего дохода (цены), эффективность рынка будет неполной. Таким образом, выбор RETLAG в качестве переменной — это один из наиболее спорных

моментов. Для каждого временного интервала мы вычисляли **невзвешенное** среднее доходов по всем зарегистрированным акциям.

Разница между ценами предложения и спроса (ASBI)

Появление этой переменной может быть оправдано (см. [39]) таким соображением: изменение величины бид-аск спреда может указывать на изменение цены акций, так как наиболее информированные спекулянты обычно сначала совершают сделки с опционами. Когда участники рынка видят, что на нем действуют более информированные лица, они увеличивают бид-аск спред. Тем самым его изменение может быть предшественником роста курса акций. На рынке одновременно существует множество бид-аск спредов, и трудно определить, который из них наиболее подходит для нашей цели. Мы для каждого временного отрезка и для каждой цены исполнения подсчитывали невзвешенное среднее по всем котировкам.

Время до исполнения (TTM)

TTM обозначает число дней, остающееся до исполнения апрельских 1992 г. опционов (которое должно произойти в четверг 16 апреля). Эта переменная была включена, поскольку она является важной входной переменной в основной формуле модели Блэка–Шоулса (там она обозначена через τ). Следуя Бодде [45], мы просто подсчитывали число остающихся дней, включая выходные и праздники (календарный подход).

Подразумеваемая волатильность (IMVOLEUR)

Эта переменная обозначает волатильность из формулы Блэка–Шоулса. Волатильность является наиболее важной экзогенной переменной этой модели, поскольку саму опционную торговлю, несколько упрощая, можно рассматривать как торговлю волатильностью. Считая, что B&S-модель OPM верна, мы можем с ее помощью определить подразумеваемую волатильность апрельских 1992 г. опционов колл всех четырех серий. Мы использовали здесь метод аппроксимации, известный как метод Ньютона–Рафсона в варианте Бенинья [37]. Цена с опциона колл есть функция величин X (цены исполнения), S (цены соответствующей акции), r (процентной ставки), σ (волатильности) и τ (времени до исполнения):

$$c = SN(d_1) - Xe^{-r\tau}N(d_2).$$

Здесь $N(\cdot)$ — функция стандартного нормального распределения,

$$d_1 = \frac{\ln(S/X) + (r + 0.5\sigma^2)\tau}{\sigma\sqrt{\tau}}, \quad d_2 = d_1 - \sigma\sqrt{\tau}.$$

Если и в самом деле рынок акций следует за рынком опционов, изменение волатильности опционов предсказывает будущую волатильность на рынке акций. Переменная IMVOMAD1 обозначает разность между подразумеваемой волатильностью и ее однодневным скользящим средним (MA) значением. Она была добавлена для того, чтобы увеличить объем исторической информации, и для того, чтобы в более явном виде сигнализировать сети о любых внезапных увеличениях или падениях подразумеваемой волатильности.

Историческая волатильность цены акций Филипс (HISVOLA)

Эта переменная, как и RETLAG, неявно проверяет свойство эффективности рынка в слабой форме. Переменная вычислялась для каждого 15-минутного интервала с помощью «движущегося окна» в 15 предшествующих дней торгов. Таким образом, историческая волатильность оценивалась каждый раз по 435 наблюдениям (15 дней $\times$ 29 интервалов) невзвешенной средней цены акций. Из-за того, что окно движущееся, первое наблюдение в обучающем множестве приходится на вторник 11 февраля 1992 г., а не на 20 января.

Процентная ставка (EURO и IMPLRE)

Процентная ставка EURO использовалась в качестве безрисковой процентной ставки (r) в формуле модели Блэка–Шоулса при вычислении подразумеваемой волатильности (IMVOLEUR, окончание EUR указывает на связь с переменной EURO). При высоких процентных ставках опционы колл становятся более привлекательными в смысле извлечения максимального дохода от инвестиций, чем операции на рынке наличности. С ростом числа инвесторов, желающих купить опционы колл, растет цена с опциона колл, и это увеличивает возможности извлечения прибыли путем продажи опционов колл и получения больших премий наличными. Чистая позиция дилеров рынка по проданным и купленным опционам колл может измениться, что, в свою очередь, может повлиять на объемы торговли и поведение цен на рынке наличности. Мы вычисляли годовую процентную ставку, соответствующую каждому сроку погашения, при помощи линейного преобразования данных ежедневного ряда Datastream по 7-дневным и 3-месячным кредитам. Поскольку промежуток времени, охватываемый исследованием, был довольно коротким (3 меся-

ца), изменения процентной ставки за это время были небольшими и происходили редко.

Нуман [206] утверждает, что ставка EURO не может служить хорошим приближением для процентной ставки в модели Блэка–Шоулса. Он показывает, что ставки AIBOR (Амстердамская ставка предложения по межбанковскому кредиту) и EURO дают разные оценки для подразумеваемой неустойчивости опционов пут и колл на одни и те же акции с одинаковыми сроками и ценами исполнения. Для решения этой проблемы Нуман предлагает использовать подразумеваемую процентную ставку (IMPLRE). Для ее вычисления нужно, чтобы в каждом 15-минутном интервале совершались сделки по апрельским at-the-money опционам пут и колл. Если это условие не выполнялось, то бралось значение с предыдущего интервала. Мы считали, что опционы определенной серии являются at-the-money, если разность между невзвешенной средней ценой акций на данном 15-минутном интервале и ценой исполнения опционов была по абсолютной величине меньше 2 флоринов.

Мы исходили из предположения, что когда рыночная ставка (EURO) больше подразумеваемой ставки на тот же срок, дилеры рынка имеют на балансе чистую короткую позицию по основным акциям, поскольку они могут продавать акции Филипс без покрытия (short sale) и затем реинвестировать полученную от продажи наличность по рыночной ставке кредита. В общем, эта переменная была включена в модель потому, что изменение подразумеваемой процентной ставки может говорить о повышении активности на рынке наличности, и потому, что она лучше, чем ставка EURO, подходит для вычисления теоретической цены опциона (THEPOP).

Аналогичным образом переменная MMPOSLO учитывает чистую короткую/длинную позицию дилеров рынка.

Формально, значение выражения $[(IMPLRE - EURO)/EURO] \times 0.5 + 0.5$, большее 0.5, указывает на чистую длинную позицию по основным акциям, а меньшее 0.5 — на чистую короткую позицию.

Теоретическая цена апрельских 1992 г. опционов колл (THEPOP)

Теоретическая цена опционов (с для краткости) четырех различных цен исполнения вычислялась с помощью модели Блэка–Шоулса определения цены опционов. Как уже было сказано, в качестве безрисковой процентной ставки мы использовали подразумеваемую ставку (IMPLRE). Далее, вместо того, чтобы работать с четырьмя подразумеваемыми волатильностями или же с одной общей, мы брали подразумеваемую волатильность σ опционов той серии, чья цена с была наиболее чувствительна к σ на данном временном интервале.

Эта чувствительность проявляется в больших значениях λ . Бекерс [34] показал, что такая волатильность наиболее точно характеризует «истинную» подразумеваемую волатильность.

Всякое отклонение действительной цены (WAVGPOP) от теоретической рассматривалось нами как возможная недо- или переоценка акций Филипс на данном временном интервале, которая, в свою очередь, может вызвать уточнение позиций на наличном рынке. Например, превышение действительной ценой опционов колл ее теоретического уровня может указывать на спрос на опционы этой серии, возможно, со стороны участников рынка, располагающих информацией, которая может воздействовать на курсы ценных бумаг.

При этом отклонения цены только тогда будут значимыми признаками пере- или недооценки опционов, когда в них учтена степень риска, которая измеряется величиной DELTAEUR. Поэтому мы включили в модель переменную IQ, большие значения которой соответствуют недооцененным опционам с относительно небольшим риском. Формально, $IQ = (THEPOP - WAVGPOP)/\delta$.

Пут-колл соотношение в течение дня (CAPUDIF)

Пут-колл соотношение говорит о настроениях рынка и показывает, в какой мере он настроен на повышение или на понижение курсов. Для каждого временного интервала мы брали разность между количествами сделок по опционам колл и пут всех серий и сроков исполнения.

В дополнение к этому переменные IOCADIF и IOPUDIF обозначают, соответственно, разницу между числом in- и out-of-the-money опционов колл и пут на данном временном интервале. Опцион колл называется in (out)-of-the-money, если для него $S \geq (<)X$. Аналогично, опцион пут будет in (out)-of-the-money, если $S \leq (>)X$.

Дельта опциона (DELTAEUR)

Величина дельта (δ) опциона показывает, насколько изменится цена опциона при малом изменении цены акции, являющейся предметом опционного контракта. Иначе говоря, дельта — это производная от цены с опциона по цене S основной акции. Величина δ вычислялась приближенным методом, исходя из стандартного нормального распределения (см. [37]). Ее значение, равное 50, соответствует at-the-money опциону, для которого вероятность того, что он будет предъявлен к исполнению, равна 50%. Малые δ соответствуют сильно out-of-the-money опционам, а близкие к единице — опционам, которые глубоко in-the-money. В ситуациях, когда неясно, в какую сторону будут развиваться события, инвесторы предпочитают

иметь позицию с нулевой дельта, т.е. полностью хеджированную. Любое изменение дельта опционов приводит к изменению дельта всей позиции, которое можно корректировать, изменяя количество опционов и основных акций в портфеле. Поскольку большинство инвесторов связаны обязательством все время иметь (дельта-) хеджированный портфель, изменение значения этой переменной может вызвать значительные изменения в объеме торговли акциями и в их цене.

Другой способ оценки чувствительности цены c опциона колл к изменению на 1 процент цены S основной акции дает так называемая эластичность опциона (OPELASEUR). Высокоэластичный опцион принесет инвестору относительно высокую прибыль, если тот правильно угадает изменение рынка. Эластичность в совокупности с числом сделок (TRANS) может нести определенную информацию о будущей цене акций. Формально, эластичность равна $\delta \times [S/c]$.

Переменная MONEY представляет собой еще одну количественную меру «торгуемости» опциона определенной серии. Она показывает, насколько данный опцион является in-the-money. Чем сильнее опцион in-the-money, тем больше затраты капитала и тем меньше заключается сделок. Трейдеры предпочитают работать с относительно недорогими опционами — at-the-money или немного out-of-the-money. Формально, $MONEY = S/X$.

Гамма опциона (GAMMAEUR)

Величина гамма (γ) измеряет, насколько изменится дельта опциона при малом изменении цены акции. Следовательно, гамма опциона — это вторая производная его цены по цене основной акции. Гамма, как правило, бывает большой для at-the-money опционов. Эта величина показывает, насколько существенное уточнение позиций по опционам происходит при изменении цены акций. Большое значение γ говорит о том, что изменение цены акций вызовет большие изменения в позициях по опционам. Такое уточнение позиций может повлиять на интенсивность торгов и на цену, поэтому изменения величины гамма имеют определенное значение для прогнозирования объема торговли и цен на рынке акций.

Лямбда (LAMBDAEUR), ρ (RHOEUR) и θ (THETAEUR) опционов

Лямбда (λ) измеряет чувствительность цены опциона к изменениям волатильности цены акции и равна производной c по волатильности акций. Участники торгов, располагающие конфиденциальной информацией, способной влиять на рыночные курсы, стара-

ются работать с опционами тех серий, у которых λ наибольшее. Расчет здесь делается на то, что последующее распространение информации будет сказываться на цене опциона в желательную сторону.

ρ (ρ) представляет собой производную цены опциона по процентной ставке r и измеряет чувствительность величины c к изменениям процентной ставки.

θ (θ) есть производная цены опциона по времени t , остающемуся до исполнения опциона. Эта величина всегда отрицательна. Опытные инвесторы обычно закрывают свои опционные позиции задолго до дня погашения опциона. Эти операции сказываются на величине открытого интереса по данной серии опционов и могут влиять на величину показателя дельта у позиций, занятых участниками торгов.

Число изменений рыночной котировки (DELQU)

Эта переменная измеряет степень нервозности участников рынка на определенном отрезке времени. Частые изменения котировок опционов с данной ценой исполнения могут говорить о том, что котировщики располагают несимметричной информацией, и это побуждает трейдеров к активности. Предполагается, что совокупность переменных TRANS, ASBI и DELQU несет в себе информацию о будущей цене акций.

ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ И ПОДГОТОВИТЕЛЬНЫЕ ТЕСТЫ

Вся совокупность данных была разбита на три множества: обучающее, подтверждающее и тестовое. Так как историческая волатильность вычислялась по принципу движущегося окна размером в 15 торговых дней, обучающее множество охватывает промежуток времени с 11 февраля (вторник) до 13 марта (пятница) 1992 г. и включает в себя 24 рабочих дня и 2784 наблюдений (29 15-минутных периодов в день и 4 цены исполнения). Из них 300 наблюдений были выделены в подтверждающее множество. Перед тем, как это сделать, мы перемешали все 2784 записи, чтобы устранить аспект, связанный с временными рядами. Для того чтобы избежать эффектов дня и недели исполнения, мы взяли для тестов промежутки с 16 марта (понедельник) по 3 апреля (пятница), и, таким образом, до исполнения (16 апреля) оставалось еще достаточно времени. На эти 15 дней торгов приходилось 1740 наблюдений.

Все 33 входных переменных были перемасштабированы так, чтобы их значения менялись от 0 до 1. Предварительно были «подрезаны» крайние значения дохода (1-процентные выбросы с обеих сто-

рон). В итоге значение 0.5 в новом масштабе соответствует нулевому доходу за соответствующий промежуток времени. Обратите внимание на то, что в проверочном множестве случаи нулевого дохода встречаются весьма часто.

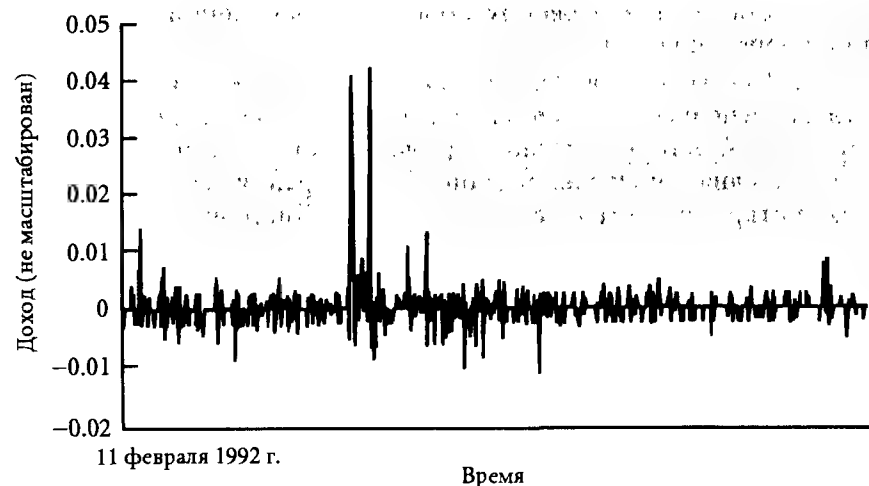


Рис. 5.1. Целевая переменная (немасштабированная): доход, полученный за четверть часа по совокупности обучающего и проверочного множеств

Для того чтобы выяснить влияние разных переменных и определить степень пригодности линейной модели, была использована линейная OLS регрессия. В силу того, что подразумеваемая ставка IMPLRE оказалась мультиколлинеарной с соотношением длинных/коротких позиций MMPOSLO (выявлено с помощью теста на допустимые отклонения пакета SPSS+/PC, версия 5.01), эта переменная была отброшена. Результаты для обучающего множества оказались весьма обнадеживающими.

Модель оказалась способной объяснять ситуацию примерно в 3% случаев — неплохой результат, когда речь идет об оценке дохода на наличном рынке на основании информации с рынка производных финансовых инструментов. Никакой корреляции ряда обнаружено не было. То обстоятельство, что на всем обучающем множестве акции Филипс медленно, но постоянно росли, учитывалось в значениях переменных TRAHOUR, HISVOLA и RETLAG.

Регрессионные данные довольно хорошо отслеживали кривую реальных доходов на первых 87 записях (что соответствует 3 торговым дням) из тестового множества (коэффициент корреляции = 13%), но на оставшемся отрезке проявлялся отчетливый тренд на понижение (большие отрицательные доходы) при коэффициенте корреляции 2%. Поскольку средний доход по акциям за 15 минут — один и тот

же для опционов всех серий, мы для каждого временного интервала вносили в график только одно значение целевой переменной, и, тем самым, число наблюдений в проверочном множестве сократилось с 1740 до 435.

Переменные	B	SE B	T	Sig T
IOCADIF	-0.1034	0.03706	-2.79	0.0053
TRAHOUR	0.02799	0.00799	3.502	0.0005
CACONOC	0.27614	0.12725	2.17	0.0301
RETLAG	0.07605	0.02011	3.782	0.0002
CACONJU	-0.0788	0.02763	-2.853	0.0044
CAPUDIF	-0.1128	0.04113	-2.742	0.0062
GAMMAEUR	-1.2795	0.49349	-2.593	0.0096
HISVOLA	0.25582	0.05204	4.916	0
EXERP	-0.336	0.06652	-5.051	0
LAMBDAEUR	0.50799	0.17884	2.84	0.0045
MONEY	-0.5069	0.16176	-3.134	0.0017
(Константа)	0.99024	0.24561	4.593	0

F статистика	3.51	Уровень значимости F = .0000
R ²	3.92%	
Уточненный R ²	2.81%	
D.W.	2.11	

Таблица 5.2. Значимые результаты регрессии на обучающем множестве

Поскольку на данных, следующих за тремя днями торгов, результаты проверки все больше расходятся с целевой переменной по величине и направленности, периодическая перенастройка модели может улучшить результаты на тестовом множестве. Не «обновляя» регрессионную модель, мы повторно обучали нейронную сеть с помощью движущегося обучающего промежутка, который охватывал 2 торговых дня или 232 записи (2 дня по 29 интервалов и 4 серии опционов). Вначале мы обучали сеть на материале обучающего (2484 записи) и подтверждающего (300 записей) множеств в течение 18,000 эпох. Затем мы делали прогноз дохода по тестовому множеству на час вперед (4 записи). После этого сеть повторно обучалась на последних 232 записях, включая те 4 интервала, для которых на предыдущем шаге был сделан прогноз. Затем делался прогноз еще на четыре 15-минутных интервала и т.д. Срок прогноза в 4 интервала и интенсивность повторного обучения в 100 эпох выбирались волевым порядком, и в дальнейшем эти параметры можно уточнить. Мы сосредоточились именно на краткосрочном (максимум на 1 час впе-

ред) прогнозе дохода, поскольку предметом всего исследования являются изменения показателей в течение одного торгового дня.

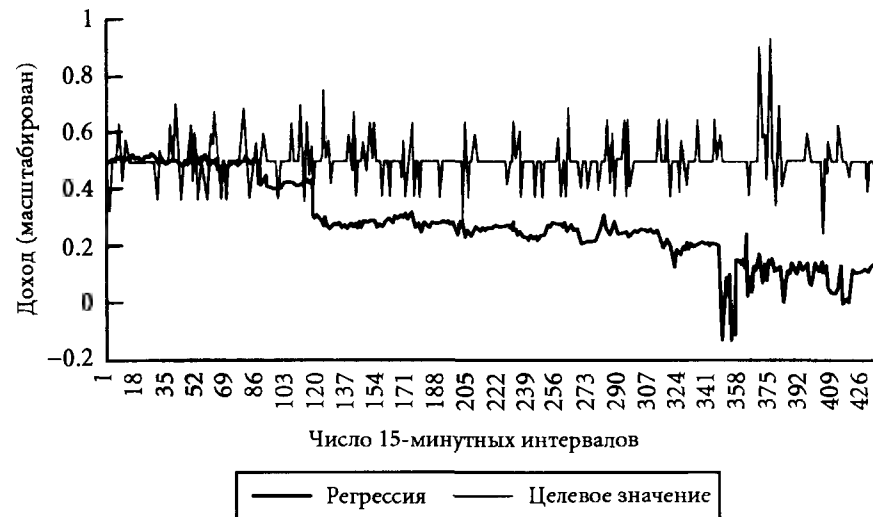


Рис. 5.2. Результаты регрессии, полученные на новых данных (тестовом множестве)

РЕЗУЛЬТАТЫ РАБОТЫ СЕТИ

С помощью 33-14-1 MBPN-сети с прямыми связями между входами и выходом и логистическими функциями активации были получены две совокупности прогнозов.

Первая совокупность прогнозов получена сетью (которую мы назвали ALLNET), обученной на исходных обучающих и подтверждающих данных в течение 18,000 эпох. Были изучены влияния разных значений коэффициента обучения, импульса и числа скрытых элементов на среднеквадратичную ошибку на подтверждающем множестве. Наименьшие значения достигались при коэффициенте обучения 0.1, импульсе 0.9 и 14 скрытых элементах.

Второе множество прогнозов было получено на выходе сети ALLNET после ее повторного обучения с помощью движущегося окна (этот вариант получил название WINNET). Ожидалось, что второй метод, по сравнению с первым и с регрессией, даст лучшие результаты при работе с новыми данными по каждому из следующих критериев:

- Средняя квадратичная ошибка (MSE) на тестовом множестве (MSETE).

- Коэффициент корреляции Пирсона ρ между целевым значением и выходом (COCO). Этот коэффициент измеряет, насколько хорошо выход совпадает с целевыми значениями в критических точках.
- Чистый доход, полученный на тестовом множестве (NETRET) при применении простейшей однопериодной стратегии торговли (без учета транзакционных издержек). Основываясь на прогнозе дохода, который дает сеть на следующий временной интервал, инвестор занимает длинную или короткую позицию по акциям Филипс и через 15 минут закрывает ее. Затем очередной прогноз сети опять укажет ему, занять ли длинную или короткую позицию, которая будет закрыта через 30 минут и т.д. Итоговый чистый результат может служить мерой точности прогноза, учитывающей направление и абсолютную величину.

ALLNET дает лучшие, по сравнению с регрессией, результаты в смысле MSE и оценки чистого дохода. Однако это не говорит о ее качественном превосходстве в прогнозировании перед линейной моделью. Дальше, чем на 3 дня торгов, и регрессия, и ALLNET прогнозировали значения меньше 0.5, т.е. отрицательные доходы, так что наш инвестор должен был бы все время держать короткую позицию.

Образец (435 записей)	ALLNET	Регрессия
MSETE	0.052	0.077
COCO	2.0%	2.3%
NETRET	-0.34%	-0.46%

Предполагая, что адаптивно обученная сеть может дать лучшие результаты, мы применили метод обучения при помощи движущегося окна. Для удобства сеть ALLNET повторно обучалась 100 раз, что дало в результате 404 прогноза (вместо 435). На выходе эта сеть (WINNET) уже не давала понижающего тренда, и это подтвердило наши представления о том, что метод адаптивного обучения имеет преимущество перед статическим обучением, использованным в ALLNET.

Как мы и ожидали, WINNET показала лучшие результаты по всем трем критериям. При этом нужно учесть, что значения пересчитаны так, чтобы учесть меньший объем тестовых данных.

Образец (404 записи)	WINNET	ALLNET	Регрессия
MSETE	0.009	0.05	0.069
COCO	1.31%	0.8%	0.46%
NETRET	1.94%	-1.11%	-1.31%

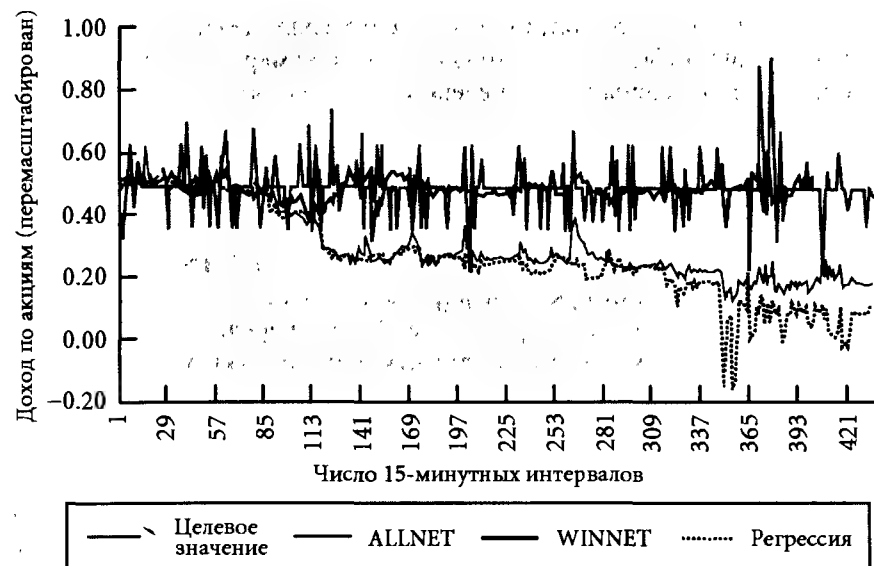


Рис. 5.3. Целевые значения (перемасштабированные) в сравнении с прогнозами сетей ALLNET и WINNET

Если говорить о всем тестовом множестве, то обе сети дают лучшие результаты, чем регрессионный анализ. Это неудивительно, поскольку сети способны улавливать нелинейности, содержащиеся в данных. Далее, адаптивная сеть (WINNET) лучше оценивает будущие доходы, чем простая (ALLNET), потому что она может прогнозировать как положительные, так и отрицательные доходы. ALLNET и регрессия дают разные результаты из-за разного числа степеней свободы, и это привносит некоторые нюансы в расстановку победителей на пьедестале почета. Так, регрессия дает несколько лучшие, по сравнению с обеими сетями, результаты для первых трех торговых дней в смысле корреляции и MSE, но не по полученному чистому доходу. Нужно помнить при этом, что качество прогноза по методу регрессии падает после первых трех дней.

ОБСУЖДЕНИЕ

При работе с набором данных о сделках, совершаемых в течение торгового дня, который был предоставлен Европейской биржей опционов (ЕОЕ), адаптивно обучаемая сеть WINNET показала лучшие результаты, чем статическая обученная сеть и регрессия, по трем выбранным критериям. Простейшая основанная на этом прогнозе торговая стратегия принесла бы доход приблизительно в 11% годовых (без учета транзакционных издержек).

Хотя эта цифра может показаться не такой большой, если сделать поправку на риск (акций Филипс), достигнутый уровень дохода свидетельствует о хороших потенциальных возможностях нейронных сетей в вопросах моделирования информации, поступающей с рынка производных финансовых инструментов.

Поскольку нейронные сети дают возможность исследовать ранее не использовавшиеся модели с нелинейной спецификацией, возможно, что их преимущества перед традиционными моделями формирования цены объясняются в большей степени недостатками спецификации последних, нежели чем свойствами эффективности рынка.

Задним числом можно сказать, что выбор акций Филипс был не самым удачным. Эти акции, будучи одним из основных предметов торгов на Амстердамской бирже, обладают большим информационным действием, и возможности для извлечения дополнительной информации невелики. После опционов на индексы курсов акций опционы на акции Филипс были наиболее торгуемым активом на ЕОЕ в 1992 году, а торговля самими акциями также активно велась на наличном рынке. Если бы были выбраны другие, менее активно торгуемые акции, то возможностей для улучшенного прогноза было бы больше, и показатель дохода в 11% годовых, вероятно, был бы превзойден.

При переходе к более простой сети за счет уменьшения количества входных переменных и серий опционов (мы считали бы разумным сосредоточиться на at-the-money опционах), безусловно, упростился бы процесс обучения, а способность сети к обобщению — повысилась.

Оценка индексов рынка акций

Эта глава посвящена вопросам определения значений макроэкономических показателей (или глобальных факторов), влияющих на доход по индексу обыкновенных акций на примере Нью-Йоркской и Амстердамской фондовых бирж. В первом случае в качестве исходных были взяты данные, собранные Ченом, Роллом и Россом [67]. Эмпирически были выбраны оптимальные архитектуры для сети, и результаты на выходе этих сетей сравнивались с оценками, полученными обычным методом регрессии по способу наименьших квадратов (OLS). Результаты у сетей оказались лучше, чем у OLS-регрессии, благодаря их способности улавливать нелинейные связи между независимыми и зависимыми переменными. Кроме того, и это более существенно, здесь будут изложены два эвристических подхода к анализу влияний отдельных переменных на поведение решения во времени.

ВЛИЯНИЕ ЭКОНОМИЧЕСКИХ ФАКТОРОВ И ПОСТРОЕНИЕ МОДЕЛЕЙ

Считается, что цены на финансовые активы подчиняются закону случайного блуждания. Так как цены реагируют на непредвиденную информацию, которая поступает дискретно во времени, то цена должна следовать схеме случайного блуждания. Более того, финансовые рынки эффективны в том смысле, что вся доступная информация немедленно отражается в ценах на финансовые активы. Это свойство эффективности называется гипотезой эффективного рынка (ЕМН) и является одним из основных положений в теории финансов (см. гл. 3). Профессора до сих пор учат студентов тому, что временные ряды для цен с большим трудом, если вообще, поддаются прогнозу, а графический анализ показателей рынка (так называемый чартизм) и технический анализ сравнивают с колдовством и астрологией.

При этом, однако, предпринятые усилия для разработки более глубоких моделей тактики размещения активов и стратегий выбора момента для сделок (см. [186], [82]) всякий раз давали результаты, опровергающие стандартную теорию. Недавние исследования свойств временных рядов для различных финансовых активов выявили присутствие в них нелинейных динамических структур. Так,

например, Ларрэн [172] установил хаотический характер поведения цен на векселя казначейства США, а Петерс [215] обнаружил хаотический аттрактор во временном ряде для индекса S&P 500. Далее, хаотическое поведение во времени легче всего описать с помощью кратковременных и больших по амплитуде возмущений, и это ставит под вопрос значимость результатов спектрального анализа и проверок на автокорреляцию (см. [51]). В то время как при анализе временных рядов линейными методами большинство видов доходов представляются результатами действия белого шума, эти авторы обнаружили, что в рядах также присутствуют, пусть небольшие, но значимые нелинейные зависимости (см. [142], [279]).

Коль скоро формально доказано наличие в динамике системы детерминированного хаотического нелинейного поведения, встает вопрос о том, какая модель за этим стоит. К сожалению, окончательного ответа здесь пока не получено (ср. [67], [148]). Для того чтобы лица, принимающие решения, считали модель подходящей, она должна быть в достаточной степени детализирована, т.е. включать в себя методы многомерного анализа меняющихся во времени, возможно, нерекуррентных зависимостей в данных. Из своего каждодневного опыта финансовый менеджер знает, что цена конкретного актива испытывает влияние множества общерыночных (систематических) и относящихся к отдельным фирмам (несистематических) факторов риска. Поскольку риски второго типа оценить трудно, они исключаются путем формирования хорошо диверсифицированного портфеля акций, и традиционная теория финансов сосредоточивается на определении систематических факторов риска, влияющих на все фирмы. В середине 60-х годов Шарп [241], развивая идеи Марковица [187], предложил модель ценообразования на рынке капиталовложений (Capital Asset Pricing Model, CAPM), которая основывалась всего на одном систематическом факторе — доходе от рыночного портфеля. Основное положение модели CAPM заключается в том, что доход от ценной бумаги на конкурентном рынке равен сумме безрискового дохода и премии за риск, которая прямо пропорциональна коэффициенту бета (β) этой бумаги, измеряющему степень рыночного риска, который не может быть устранен диверсификацией:

$$\bar{R}_{i,t} = R_f + \beta (\bar{R}_{m,t} - R_f), \quad (1)$$

где $\bar{R}_{i,t}$ — ожидаемый доход по i -й ценной бумаге в момент t , R_f — доход по безрисковым ценным бумагам, $\bar{R}_{m,t}$ — ожидаемый доход от рыночного портфеля (m) в момент t .

Благодаря своей простоте, модель CAPM пришла ко двору в сообществе инвесторов. При этом, однако, остались некоторые проблемы. Во-первых, модель, в принципе, не может быть полностью

определена, потому что ожидаемый уровень дохода невозможно определить по текущим доходам. Тот доход, который по прошествии определенного времени, действительно, удастся получить, не обязательно совпадет с тем, что ожидается. Далее, рыночный портфель должен включать в себя все возможные виды рисковых инвестиций — облигации, акции, товары, недвижимость, предметы искусства, вложения в человека. На деле вместо всего этого берется какой-то из индексов акций (например, FTSE 100 или S&P 500). Что еще хуже, модель не согласуется с фактами. Правомерность использования показателя β стала вызывать сомнения после того, как выяснилось, что он практически не объясняет уровень фактически полученного дохода (см. [63]).

Все это вызвало появление альтернативных теорий ценообразования активов. Предложенная Россом [233] теория арбитражного оценивания (Arbitrage Pricing Theory, APT) утверждает, что ожидаемая премия за риск для некоторой акции определяется ожидаемыми премиями за риск, соответствующими различным факторам, и чувствительностью акции к каждому из этих факторов. Одним из таких факторов может быть доход от рыночного портфеля, так что модель CAPM можно рассматривать как частный случай модели APT. Таким образом, модель APT учитывает различные источники систематического риска:

$$\bar{R}_{i,t} = R_f + \sum_{x=1}^n b_x (\bar{R}_{\text{factor } x} - R_f), \quad (2)$$

где $\bar{R}_{i,t}$ — ожидаемый доход по i -й ценной бумаге в момент t , R_f — доход по безрисковым ценным бумагам, x — совокупность макроэкономических факторов, b_x — чувствительность дохода по i -й ценной бумаге к изменениям фактора x , $\bar{R}_{\text{factor } x}$ — ожидаемый доход от макроэкономического фактора x .

Из экспериментов видно, что APT намного превосходит CAPM при применении к индексу S&P 500 и другим взвешенным по капиталу или равновзвешенным индексам (см. [66], [193]).

Главная трудность при тестировании модели APT связана с определением количества факторов и соответствующих им коэффициентов (см. [267]). Хотя в экономической теории рассматривается большое число глобальных экономических факторов, для объяснения изменений уровня дохода обычно бывает достаточно пяти из них.

ЛИНЕЙНАЯ МОДЕЛЬ АРТ

В известной работе Чена, Ролла и Росса [67] делается вывод о том, что доход по акциям практически линейно зависит от изменений следующих пяти макроэкономических факторов: разницы между долгосрочными и краткосрочными процентными ставками, предвиденной инфляции, непредвиденной инфляции, промышленного производства и различия между первоклассными и прочими облигациями. База данных, с которой работали авторы, включает в себя 396 ежемесячных наблюдений за период с 1953 г. по 1985 г. В аналитической записи:

$$R_{VWNY,t} = a + b_{MP} MP + b_{DEI} DEI + b_{UI} UI + b_{URP} URP + b_{UTS} UTS + e, \quad (3)$$

где b — влияния макроэкономических переменных, a — константа, e — погрешность, или специфический риск данной компании.

Несомненно, кандидатами на включение в модель являются и другие макроэкономические переменные. Чен, Ролл и Росс [67] рассматривали потребление товаров и услуг, а также цену на нефть. Оказалось, что эти факторы существенно не влияют на ожидаемый доход, и поэтому они были исключены из дальнейшего анализа. В связи с ростом международной торговли Кимото и др. [162] предложили учитывать курсы обмена валют и индекс Доу-Джонса для прогнозирования индекса Токийской фондовой биржи. Определенную дополнительную информацию может дать показатель предложения денег (см. [23]). Далее, может оказаться полезным учет показателей рынка производных финансовых инструментов (опционов и фьючерсов), поскольку они отражают ожидаемые события в экономической жизни. Разумеется, при этом нужно решить вопрос о том, который из рынков — наличности или рынков производных финансовых инструментов — опережает другой во времени.

Чен, Ролл и Росс разбивали всю совокупность данных на три подынтервала и оценивали значимость каждого из шести факторов во времени при помощи множественного регрессионного анализа по методу наименьших квадратов.

Оказалось, что месячное производство и непредвиденная инфляция существенной роли не играют, а премия за риск и временная структура всегда имеют значительное влияние. Показатель годового производства играет роль только на больших промежутках времени. Изменения ожидаемой инфляции стали значимым фактором только в самое последнее время, возможно, это — результат монетаристской политики администрации Рейгана.

$U[I(t)]$	Непредвиденная инфляция	$I(t) - E[I(t) t-1]$ $I(t)$ = логарифм индекса цен на потребительские товары США, $E[I(t)]$ = ожидаемая инфляция (см. [108])
$DEI(t)$	Изменение ожидаемой инфляции	$E[I(t+1) t] - E[I(t) t-1]$
$URP(t)$	Премия за риск	$Baa(t) - LGB(t)$ $Baa(t)$ = доход по облигациям класса Baa и ниже (за 1953–77 гг. — данные Ibbotson, 1979; 1978–83 гг. — данные авторов) $LGB(t)$ = доход по долгосрочным государственным облигациям (1958–78 гг. — данные Ibbotson and Sinquefeld, 1981; 1979–83 гг. — CRSP)
$UTS(t)$	Временная структура	$LGB(t) - TB(t-1)$, $TB(t)$ — доход по казначейским векселям
$MP(t)$	Месячный рост промышленного производства	$\ln[IP(t)/IP(t-1)]$
$YP(t)$	Годовой рост промышленного производства	$\ln[IP(t)/IP(t-12)]$
$VWNY$	Взвешенные акции	доход по равновзвешенному портфелю акций, котируемых на NYSE (CRSP)

Таблица 6.1. Обозначения и определения переменных

Рассматривая поведение переменных на разных подынтервалах, можно заметить, что некоторые из них не оказывают никакого влияния — они не значимы или неактивны, а другие активны на всем промежутке времени или на некоторой его части. Переменные, которые активны только на части промежутка времени, бывает труднее всего распознать при помощи методов типа OLS-регрессии, когда минимизируется квадратичная ошибка на всем интервале. Тем самым, от длины набора данных зависит, является ли некоторая переменная (например, DEI) активной или нет.

		$YP(t+12)$	$UI(t)$	$UTS(t)$	$MP(t+1)$	$DEI(t)$	$URP(t)$	Кон- станта
1953–1985	Beta	0.135	-0.063	0.592	-0.042	0.077	0.524	
	T	3.027	-1.447	9.847	-0.942	1.627	9.222	3.142
	Sig T	0.003	0.149	0.000	0.0347	0.105	0.000	0.0002
	R^2 уточн.	0.27						
	D.W.	2						
1953–1967	Beta	0.155	0.026	0.222	-0.056	-0.092	0.517	
	T	2.275	0.382	2.592	-0.812	-1.349	5.889	2.411
	Sig T	0.024	0.703	0.010	0.418	0.179	0.000	0.017
	R^2 уточн.	0.22						
	D.W.	2						
1968–1977	Beta	0.099	-0.150	0.590	-0.067	0.027	0.533	
	T	1.240	-1.954	6.016	-0.852	0.325	5.698	0.531
	Sig T	0.218	0.053	0.000	0.396	0.746	0.000	0.596
	R^2 уточн.	0.35						
	D.W.	1.9						
1978–1985	Beta	0.090	-0.140	0.985	-0.024	0.372	0.576	
	T	1.012	-1.518	6.877	-0.275	3.612	4.365	3.034
	Sig T	0.314	0.133	0.000	0.784	0.001	0.000	0.003
	R^2 уточн.	0.34						
	D.W.	1.9						
Обучающе-множественно жест-во	Beta	0.124	-0.056	0.460	-0.036	0.056	0.507	
	T	2.498	-1.158	7.280	-0.723	1.058	8.612	-2.769
	Sig T	0.013	0.248	0.000	0.470	0.291	0.000	0.006
	R^2 уточн.	0.24						
	D.W.	1.9						
1953–1980								

Таблица 6.2. Результаты регрессии для VWNY

МНОГОСЛОЙНАЯ СХЕМА С ОБРАТНЫМ РАСПРОСТРАНЕНИЕМ ОШИБКИ

С применением алгоритма Nevada Quickrop (см. гл. 1) на мэйнфрейм-машине Convex были опробованы сети различной архитектуры. Базовая структура сети проста: входной вектор из шести переменных и одномерный выход (переменная VWNY). Остается выбрать число скрытых слоев и число нейронов. Далее, поскольку результаты Чена, Ролла и Росса указывают на присутствие линейных связей между входами и выходом, способность сети к обобщению может быть увеличена за счет прямых связей между входными и выходными элементами.

В отсутствие каких-либо готовых схем для оптимального выбора модели исследователь должен опробовать различные статистические критерии согласия. Так, Утанс и Муди [270] оценивали риск предсказания, полученный при различных архитектурах сети, а Каяма и др. [157] находили общее число дублирующих друг друга элементов в скрытом слое. Мы же просто сравнивали величины квадратного корня из среднеквадратичной ошибки (RMSE) на тестовом множестве, состоящем из 60 наблюдений, относящихся к последним 5 годам интервала наблюдений (1981–85 гг.). Для дальнейшей работы была взята та архитектура сети, которая давала наименьшее RMSE.

Конфигурация	Прямые связи	Эпохи	(RMSE)
Коэффициент обучения 0.9			
6-2-1	нет	29000	0.12014
6-3-1	нет	6500	0.10687
6-4-1	нет	3000	0.10881
6-5-1	нет	6000	0.10872
6-5-4-1	нет	1500	0.10872
6-2-1	есть	7000	0.10819
6-3-1	есть	13000	0.10468
6-5-1	есть	1500	0.10918
Коэффициент обучения 2			
6-3-1	нет	4500	0.10739
6-3-1	есть	22000	0.10529

Таблица 6.3. Квадратный корень из средней квадратичной ошибки на подтверждающем множестве для полнотью обученных сетей различной архитектуры

Для того чтобы ошибочно не принять раньше времени локальный минимум погрешности обобщения за глобальный, наш алгоритм брал вдвое большее число эпох по сравнению с тем, на котором достигалось наилучшее обобщение. Таким образом, на самом деле, число эпох было вдвое больше, чем показано в табл. 6.3 и на рис. 6.1. При любом выборе коэффициента обучения ошибка RMSE на тестовом множестве оказывалась меньше, чем на обучающем. Этот в некоторой степени удивительный эффект может объясняться наличием белого шума в обучающем множестве и его отсутствием в тестовом множестве. Поскольку обучение прекращалось, как только RMSE на тестовом множестве начинала расти, мы полагаем, что переобучение не имело места, и что сеть не запоминала шум. Таким образом, относительно большая погрешность на обучающем множестве объясняется именно белым шумом.

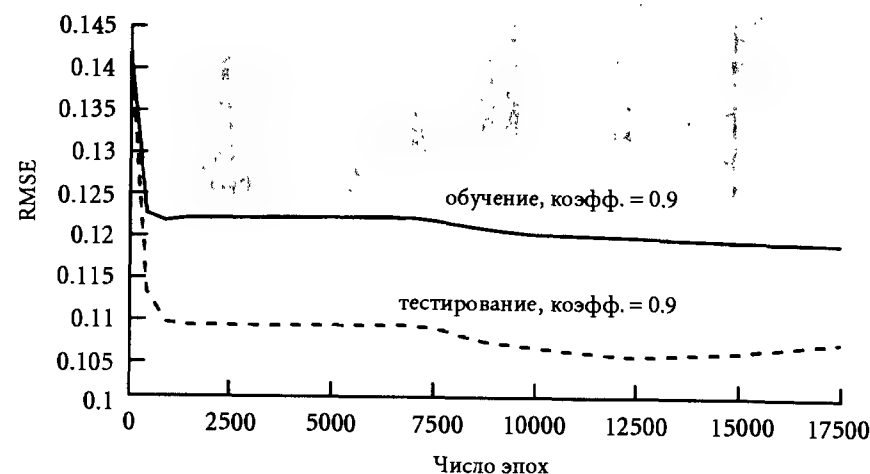


Рис. 6.1. Квадратный корень из среднеквадратичной ошибки для 6-3-1 сети с прямыми связями и коэффициентом обучения 0.9

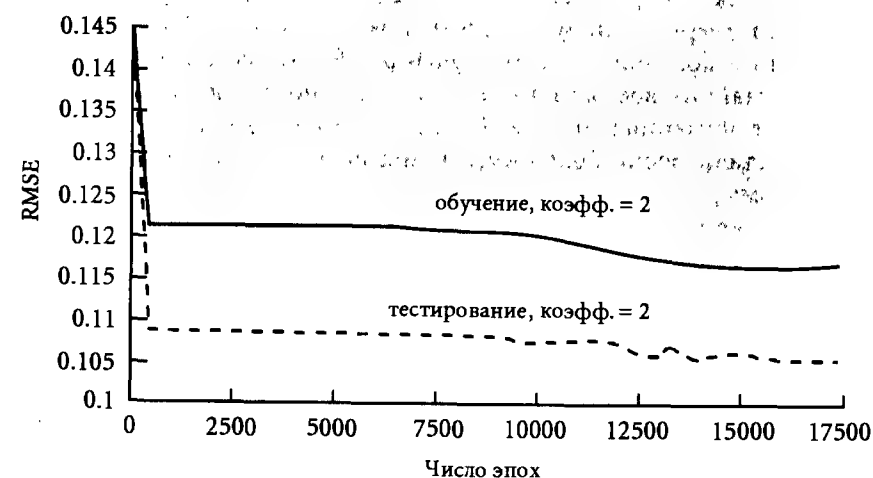


Рис. 6.2. Квадратный корень из среднеквадратичной ошибки для 6-3-1 сети с прямыми связями и коэффициентом обучения 2

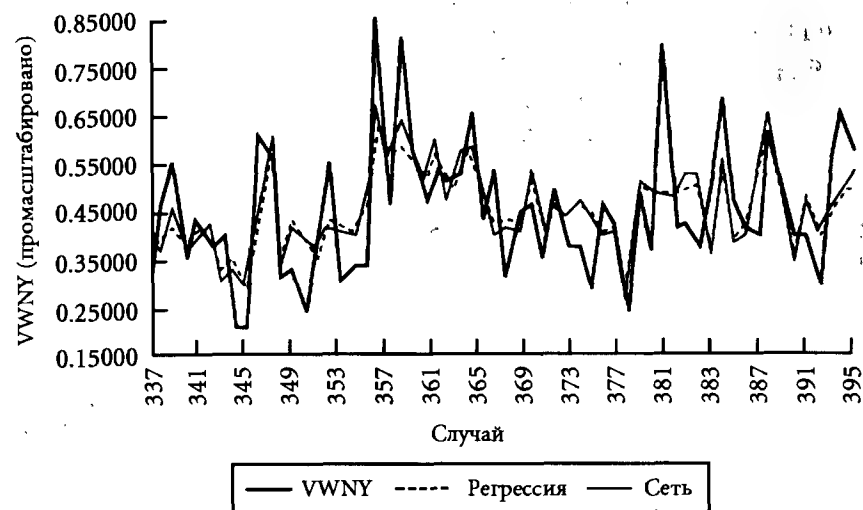


Рис. 6.3. Сравнение оценочных значений переменной VWNY, полученных регрессией и сетью, с ее истинными значениями

Среди всех конфигураций наилучшей (имеющей наименьшее RMSE на подтверждающем множестве) оказалась 6-3-1 сеть с прямыми связями и коэффициентом обучения 0.9. Желая получить решение за кратчайшее время (в пределах 13 тыс. эпох), мы увеличили коэффициент обучения в два раза (до 2). Шаги в направлении градиента теперь стали очень большими, и мы перескакивали через решение. Поэтому даже оптимально обученной сети понадобилось гораздо больше, чем 13 тыс. эпох (а именно, 22 тыс.). На рис. 6.1 видно, как RMSE быстро убывает в первые 500 эпох, а после 12 тыс. эпох начинает осциллировать.

На рис. 6.3 показаны оценки, полученные на подтверждающем множестве (которое соответствует 1981–85 гг.) с помощью OLS-регрессии и 6-3-1 сети.

Оценки, которые выдала сеть, оказались лучше всех, которые дает регрессионный анализ, как по показателю RMSE, так и коэффициентам корреляции Пирсона. При этом результаты, которые сеть показывает на новых образцах, даже превосходят те, которые регрессия имела на обучающем множестве (REG1).

До сих пор мы сравнивали между собой сетевые архитектуры с различным числом скрытых слоев и нейронов, предполагая, что каждый входной сигнал, действительно, влияет на результат. Однако, как уже говорилось, непредвиденная инфляция (UI) и месячное производство (MP) существенно не влияют на среднеквадратичную ошибку. В связи с этим возникает вопрос о том, нельзя ли эти переменные безболезненно изъять из дальнейшего рассмотрения. Явля-

ются ли эти переменные просто несущественными, т.е. не влияющими на выходной сигнал, или же они избыточные, т.е. могут быть представлены в виде линейной (или нелинейной) комбинации остальных?

RMSE на обучающем и тестовом множествах

	REG1	REG2	Сеть
обучение	0.1207	0.123	0.1191
тестирование	—	0.109	0.1047

Корреляция Пирсона между проверочными множествами:

	VWNY	Регрессия	Сеть
VWNY	1		
Регрессия	.66**	1	
Сеть	.68**	.97**	1

** — при 1-процентных хвостах распределения

Таблица 6.4. Критерии согласия для выходов регрессии и сети

RMSE

	DC631	UI(t)	DEI(t)	URP(t)	UTS(t)	MP(t+1)	YP(t+12)
Общ.	0.117	0.118	0.120	0.141	0.148	0.117	0.119
Обуч.	0.119	0.120	0.123	0.143	0.139	0.119	0.121
Тест.	0.105	0.106	0.108	0.128	0.191	0.106	0.108
Разница в RMSE (в процентах)							
Общ.	100	0.94	2.88	20.30	26.45	0.33	1.74
Обуч.	100	0.91	2.87	20.02	16.54	0.26	1.51
Тест.	100	1.19	2.95	22.27	82.91	0.82	3.39

Таблица 6.5. Вклад всех переменных в решение на обучающем, тестовом множествах и на всех данных

Чтобы оценить вклад переменных, мы вычисляли выход сети с оптимальным вектором весов с помощью пакета Microsoft Excel, версия 4. Затем про каждую переменную по очереди мы временно полагали, что ее значение неизвестно и должно быть заменено на среднее арифметическое (безусловное ожидание) этих значений при постоянных исходных значениях остальных пяти переменных. В результате получилось 6 новых входных матриц. Затем мы вычислили выходы сети для всех этих матриц. Для шести полученных выходных рядов подсчитывалась RMSE и сравнивалась с RMSE исходной входной матрицы. Идея была в том, что для переменной, которая активно влияет на решение, RMSE на соответствующем выходном векторе должна быть заметно больше, чем для исходной входной матрицы.

Все вычисленные таким образом RMSE оказались больше исходной. Такое увеличение означает, что замена переменной ее безуслов-

ным ожиданием ухудшает оценку целевой переменной. В случаях с временной структурой и премией за риск рост RMSE был самым большим (соответственно, 83% и 22% на проверочном множестве). Вспомните, что эти две переменные по результатам регрессионного анализа на всех подинтервалах также были оценены как имеющие сильное влияние. Далее, выявилось такое любопытное обстоятельство: переменные, которые по результатам регрессионного анализа были квалифицированы как неактивные, на самом деле, влияют на решение. Непредвиденная инфляция и месячная продукция имеют определенную «объясняющую роль» и не могут быть заменены комбинациями других входных переменных.

Различия между сетью и OLS-регрессией становятся разительными, когда в данных присутствуют нелинейности, которые можно уловить с помощью сигмоидальной функции преобразования. Вигенд [275] ввел следующую меру улучшения результата за счет функции активации по сравнению с OLS:

$$\mathcal{R} = \frac{\text{дисперсия остатков (нелинейная модель)}}{\text{дисперсия остатков (линейная модель)}}. \quad (4)$$

Ее значение всегда лежит в интервале от 0 до 1, поскольку от того, что сеть при обучении улавливает содержащиеся в данных нелинейности, погрешность может только уменьшиться. Значения этого отношения для обучающего и проверочного множеств оказались равны, соответственно, 0.94 и 0.92, и это говорит о том, что либо сеть плохо использует свои нелинейные возможности, либо нелинейностей в данных просто нет. Мы подозреваем второе, потому что база данных строилась с помощью линейных моделей, для того чтобы выделить взаимно не коррелирующие экономические факторы. Большим значением данного отношения объясняется то обстоятельство, что обученная сеть лишь незначительно превосходит OLS-регрессию по критерию RMSE. Однако остается фактом то, что нейронные сети превосходят OLS-регрессию даже при работе с такими данными, в которых нелинейные связи между входами и целевой переменной выражены слабо.

СРАВНЕНИЕ ИНДИВИДУАЛЬНОГО И СИСТЕМАТИЧЕСКОГО ВКЛАДА ПЕРЕМЕННЫХ

Достаточно широко распространено мнение, что нейронные сети не дают ничего для понимания пользователем сути проблемы и, в отличие от регрессионного анализа, который выдает ряд статистик по каждой переменной, остаются «черным ящиком». В этой связи в данном разделе мы приводим результаты исследования факторов, влияющих на решение, которое принимает сеть. В идеале следовало

бы выяснить значимость либо незначимость отдельного вклада каждой из переменных и предоставить лицу, принимающему решение, возможность самому разбивать переменные по категориям в соответствии с их поведением во времени при различных обстоятельствах.

Представим себе, например, переменную, которая влияет на выход (доход по индексу курсов акций) только на одной из фаз делового цикла (скажем, при понижении конъюнктуры рынка, но не при подъеме, или же наоборот). Общий (усредненный) ее вклад может быть незначительным, и OLS-регрессия, скорее всего, квалифицирует эту переменную как несущественную. Теперь представим себе, что некоторая переменная активна во всех фазах цикла, но в разные фазы действует в разных направлениях (например, переменная, усиливающая тенденцию рынка и на его подъеме, и на спаде). Такое типично нелинейное поведение может остаться совершенно незамеченным OLS-регрессией, но MBPN-сеть, скорее всего, его уловит. Далее, предположим, что переменная активна во всех фазах делового цикла и действует в том же направлении, что и движется рынок (например, увеличивает доходы во время роста активности на рынке и уменьшает их при понижении рынка). Вероятно, и сеть, и регрессия скажут, что эта переменная значима. Итак, классификация переменных в соответствии с их вкладом (поведением) во времени может пролить свет на механизмы происхождения дохода от акций.

Коль скоро OLS-регрессия не всегда способна уловить все имеющиеся функциональные связи между независимыми и зависимыми переменными, нужно искать другие пути к пониманию поведения переменных. Мы расскажем здесь о двух интуитивных эвристических подходах. В первом из них важность переменной оценивается путем сравнения погрешности прогноза, полученного при исходной входной матрице, с погрешностью, которая получится, если значения всех переменных заменить на их средние значения. Во втором эвристическом методе вклад отдельной переменной оценивается по степени надежности выхода сети (decisiveness). Метод работает «наперед» (ex ante), не обращаясь к реальным значениям целевой переменной или погрешности. Его недостаток состоит в том, что переменные могут быть классифицированы в соответствии с тем, поддерживают или противоречат ли они выдаваемому решению, а это решение на самом деле может быть неправильным.

Анализ поведения переменных на основе величины погрешности

Изучив отклонения выхода сети от целевой переменной (VWNY) для шести различных входных матриц, описанных выше, можно сде-

дать определенные выводы об относительных изменениях влияний отдельных переменных.

Дата	HIT/MISS	UI(t)	DEI(t)	URP(t)	UTS(t)	MP(t+1)	YP(t+12)
840131	-31	-81	-79	-100	-25	-93	-77
840229	-47	-62	-60	-27	-100	-61	-59
840330	25	80	100	62	19	70	73
840430	8	39	45	100	-100	40	40
840531	-18	-20	-27	-8	-100	-25	-23
840629	-7	-80	-100	100	57	-43	-87
840731	-45	-46	-47	-100	100	-49	-48
840831	100	82	81	78	100	81	81
840928	-18	-49	-55	-100	100	-44	-50
841031	-32	-37	-65	-100	100	-56	-62
841130	-47	-99	-87	-67	-84	-91	-100
841231	45	95	95	40	100	92	94
850131	39	50	50	57	100	46	48
850228	30	33	56	100	-100	48	43
850329	3	-1	-1	-100	100	16	-3
850430	-35	-92	-87	-86	-31	-93	-100
850531	-16	-51	-52	-100	100	-51	-57
850628	-13	-80	-94	100	-10	-48	-100
850731	20	100	89	97	-100	81	88
850830	-23	-70	-76	-100	100	-73	-79
850930	-37	-83	-81	-83	-100	-77	-88
851031	34	50	52	7	100	53	47
851129	47	53	50	29	100	56	50
851231	13	17	24	-100	100	16	14

Таблица 6.6. Вклад отдельных переменных в погрешность (промасштабирован на интервал [-100,100]) за период 1984–85 гг.

В столбце, обозначенном HIT/MISS, приведены отклонения от целевого значения того прогноза, который 6-3-1 сеть сделала по исходной входной матрице. Погрешность всюду была промасштабирована так, чтобы значения располагались от -100 до 100, при этом положительный знак соответствует превышению цели, а отрицательный — недобору. Малые по абсолютной величине числа означают точный прогноз (например, апрель 1984), а большие — значительную ошибку (август 1984). Если абсолютная величина ошибки велика, скажем, больше 40, то в этом случае определить вклад отдельной переменной затруднительно. Следующие 6 столбцов таблицы содержат отклонения выхода сети от целевого значения, соответствующие шести описанным выше входным матрицам. Здесь погрешности также промасштабированы и лежат от -100 до 100. По этим данным

уже можно судить о динамике отдельных переменных. Можно заметить, что во всех случаях величина погрешности чувствительна к изменениям значений переменных, и это говорит о том, что все переменные активны. В целом, по-видимому, наибольшие погрешности связаны с переменными временной структуры и премии за риск. Замена истинных значений этих переменных их средними значениями дает относительно большую ошибку прогноза. Это замечание согласуется с результатами регрессии, согласно которым обе переменные являлись высокосвязными. Если теперь мы обратим внимание на знак погрешности, то увидим здесь для этих переменных противоположную зависимость. Большие отрицательные погрешности для одной переменной, как правило, совпадают с большими положительными для другой, и это может указывать на то, что их вклады в погрешность в некоторой степени компенсируют друг друга. Поскольку все переменные прошли проверку на допустимость, мультиколлинеарность может быть исключена (см. [115]). Что касается переменных, которые по результатам регрессии были отнесены к незначимым, — например, непредвиденная инфляция, — то они, в основном, менее активны и редко дают абсолютные значения 100. Несмотря на то, что такие переменные, как показывает регрессионный анализ, в среднем мало влияют на результат, в некоторых ситуациях они могут быть очень активными. Поскольку такая маргинальная активность не улавливается регрессией, исследование чувствительности погрешности к изменениям этих якобы малозначимых переменных может продвинуть нас в понимании того, какое влияние глобальные факторы оказывают на рынок акций.

Анализ влияния переменных, основанный на решающем правиле классификации

Теперь мы рассмотрим другой метод решения задачи, который основан на решающем критерии классификации. Этот критерий заложен в разработанный Хехт-Нильсеном программный пакет KnowledgeNet для MBPN-сетей, предназначенный для принятия одного или нескольких бинарных решений.

Бинарный выходной узел сети выдает один из двух сигналов — TRUE или FALSE (ИСТИНА или ЛОЖЬ). Такая схема, скорее, может быть использована в сети, предназначенной для классификации периодических доходов, а не для прогнозирования. Каждому классу доходов соответствует один выходной узел. Мы выделили 4 класса доходов по индексу: 0 — очень низкий (т.е. сильно отрицательный), 1 — умеренно низкий, 2 — умеренно высокий, 3 — сильно положительный. Границы классов были установлены так, чтобы во все классы попадало примерно поровну случаев. Для описания четырех клас-

сов необходимы два бинарных элемента, каждый из которых дает на выходе либо 0 (FALSE), либо 1 (TRUE). Таким образом, возможны 4 выходных комбинации: 00, 01, 10, 11. Каждый из выходных элементов порождает внутренний сигнал с значениями от 0 до 1, а решающее правило с порогом, установленным на 0.5, квалифицирует сигналы, превышающие 0.5, как TRUE, а меньшие 0.5, — как FALSE. Сеть была обучена проводить классификацию дохода по индексу путем последовательной подачи на ее вход обучающего сигнала, состоящего из показателей действительно полученных доходов.

После того, как сеть обучена, становится возможным проследить для каждого входного решающего (determinant) вектора результаты классификации реальных доходов. Для каждого временного отрезка мы можем вычислить так называемое решающее значение классификации. Эта величина показывает, насколько оба внутренних выходных сигнала были далеки от порога, установленного для принятия решения («сила сигнала»). В нашей реализации мы просто берем среднее сил обоих сигналов. Расстояние от сигнала до порога может принимать значения от 0 до 0.5. Мы берем 0.5 за 100%, так что величина решающего значения может меняться от 0 до 100. Теперь можно определить вклад каждой из компонент входного вектора в решающую способность на взятом отрезке времени. Делается это так: временно предполагается, что значение компоненты неизвестно, и изучается изменение решающей способности на выходе. Вместо неизвестного входного значения внутрь нейронной сети вводится среднее арифметическое значение (или безусловное математическое ожидание) соответствующих входных значений. После того, как влияние всех входов вычислены, они масштабируются так, чтобы наибольшая абсолютная величина вклада у каждого входа равнялась 100. В табл. 6.7 представлен репрезентативный временной срез выходных значений.

Столбец «Выход» содержит результаты классификации, выданные сетью, а в столбце «Цель» указан настоящий номер класса. «ABSERR» — это абсолютная ошибка классификации, т.е. расстояние до настоящего класса, а «DECISIVN» — определенная выше величина решающей способности. На всем материале не произошло ни одной грубой ошибки в классификации — величина ABSERR ни разу не превосходит двух. Переменная временной структуры имеет сильную распознающую роль. Премия за риск, наоборот, представляется лишней переменной. Месячное производство приобрело несколько большее значение, в то время как вклад переменных, выражающих инфляцию, неясен. Результаты такого анализа, которые репрезентативны для всего набора данных, не вполне согласуются с той интерпретацией роли переменных, которую мы получили при изучении погрешности. Однако это противоречие — скорее, кажущееся, по-

скольку KNET разбивает наблюдения на классы без учета величины и знака реального дохода.

Дата	Цель	Выход	ABSERR	DECISIVN	UI	DEI	URP	UTS	MP	YP
810130	0	1	1	9	-9	-2	7	100	35	-2
810227	2	1	1	13	2	-1	1	100	55	-1
810331	3	1	2	4	-57	90	-13	31	100	-14
810430	1	1	0	66	-10	-1	1	100	59	-1
810529	2	1	1	17	-22	-21	18	-26	100	2
810630	1	1	0	2	-11	4	-6	100	48	1
810731	2	1	1	5	1	-3	-7	100	90	-1
810831	0	1	1	4	-3	-3	0	100	72	-1
810930	0	0	0	34	1	70	-12	100	-44	0
811030	3	2	1	3	45	-3	11	62	100	0
811130	3	3	0	87	3	1	4	100	-8	0
811231	1	1	0	95	-1	1	2	100	79	3
820129	1	1	0	43	5	-7	15	100	9	-13

Таблица 6.7. Классификация вкладов переменных в фактор решающей способности (KNET)

Анализ маргинального влияния переменных имеет смысл только тогда, когда коэффициент решаемости DECISIVN велик, а фактическая ошибка классификации ABSERR (равная разности номеров действительного и спрогнозированного классов) мала. Иначе говоря, если классификация оказалась успешной, можно ожидать сильную обратную зависимость между решающей способностью и ошибкой классификации. На рис. 6.4 показано совместное распределение решающей способности (упорядочена по убыванию) и абсолютной ошибки. Хорошо видно, что число случаев неправильной классификации и величина ошибки растут с убыванием решающей способности — что и следовало ожидать.

Как уже говорилось, набор переменных, предложенный Ченом, Роллом и Россом, может быть расширен с тем, чтобы учитывалась международная торговля. Мы попытались выяснить зависимости между так называемыми глобальными факторами и месячным доходом по Амстердамскому индексу обыкновенных акций Datastream за период с ноября 1979 г. по март 1991 г. (см. [23]). Для того чтобы учесть особенности голландского рынка, было добавлено 11 новых макроэкономических переменных (см. табл. 6.8), в том числе: предложение денег (M2), средний курс обмена доллар/гульден (FUS), условия торговли (TERMS).

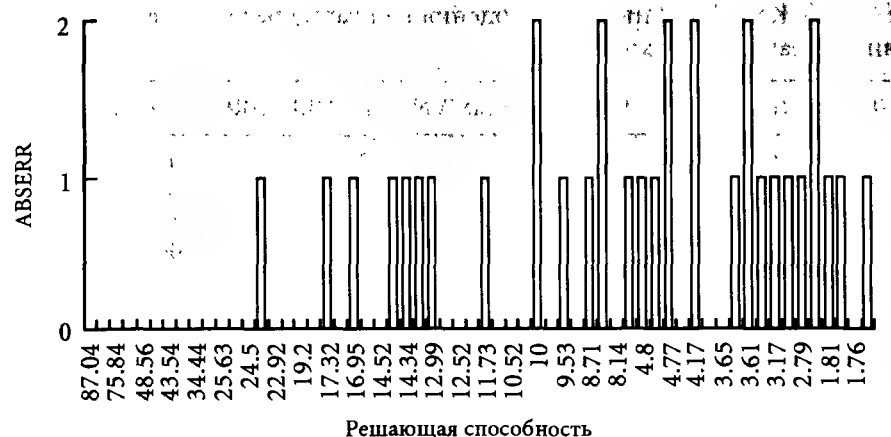


Рис. 6.4. Случаи неправильной классификации сетью (упорядочены)

Обозначение	Описание
1. CALLRATE	ставка по ссудам до востребования
2. LTGOVB	проценты по долгосрочным государственным облигациям
3. M2	предложение денег
4. TBNLUS	разница между процентными ставками по векселям казначейств Нидерландов и США
5. TESTRU	временная структура (LTGOVB минус ставка по казначейским векселям)
6. ECPI	предполагаемый процент инфляции
7. UCPI	непредвиденная инфляция
8. DECPI	изменения предполагаемой инфляции
9. RISKP	премия за риск
10. FUS	средний курс обмена гульден/доллар
11. TERMS	условия торговли (отношение экспортных цен к импортным, в том числе BLEU & Oil)
12. NONDUG	изменения потребления товаров недлительного пользования
13. INDPRO	валовое промышленное производство
14. ORDANO	экспертная оценка активности рынка: число распоряжений о сделках, расцененных как превышающие обычный уровень
15. ORDBNO	число распоряжений о сделках ниже обычного уровня
16. RETLAG	доход по индексу акций за предыдущий период времени
17. LEVELLAG	уровень индекса акций в предыдущий период
18. RETIN (целевая)	доход по Амстердамскому индексу акций (по данным Datastream)

Таблица 6.8. Обозначения и описания переменных

Архитектура сети такова: 17-мерный входной вектор, один скрытый слой из 9 элементов, и все эти узлы имеют непосредственные соединения с двумя бинарными элементами выходного слоя. В табл. 6.9 приведены репрезентативные результаты классификации и влияния отдельных переменных.

В последних 17 столбцах показан вклад каждой переменной в решающую способность. Рассмотрим, например, классификацию на февраль 1989 г. (вторая строка сверху). Класс указан правильно (большой положительный доход) и решающая величина большая. Предложение денег, потребление товаров краткосрочного пользования и доход по долгосрочным государственным облигациям вносят наибольший вклад в решающую способность в том смысле, что они отчетливо подтверждают результат классификации. Наоборот, премия за риск и обменный курс валюты резко противоречат результату классификации.

Вклад переменной RETLAG (доход за время с $t-2$ до $t-1$) невелик, и этого следовало ожидать, поскольку изменения зависимой переменной описываются если не случайным блужданием, то, во всяком случае, мартингалным процессом (см. [177]). Существенная роль переменной LEVELLAG (предыдущий показатель индекса) подтверждает гипотезу о том, что уровень индекса является важным лидирующим показателем для ожидаемого дохода. То обстоятельство, что регрессия не квалифицирует эту переменную как значимую, может говорить о том, что в механизме влияния уровня индекса на доход имеется какая-то асимметрия, возможно, связанная с нелинейной обратной связью. Это наблюдение интуитивно придает дополнительный вес гипотезе встречных инвестиций, которую высказал Чан [62] и ряд других авторов. Поведение рынка, для которого характерны возвращения к средним значениям, если оно, действительно, имеет место, неявно опровергает гипотезу эффективного рынка.

ВЫВОДЫ

В отличие от формальной модели CAPM, модель ART позволяет строить интуитивную стратегию управления риском от дохода по индексу путем выбора такой совокупности факторов, что несистематический риск для каждой ценной бумаги становится некоррелированным с несистематическим риском по любой другой бумаге. Нейронные сети имеют преимущество перед более традиционными методами в случаях, когда мы не можем точно описать все имеющиеся взаимосвязи, но можем выделить некоторый набор показателей, характеризующий исследуемое явление. В отсутствие четкой концептуальной модели регрессионные методы неприменимы.

Случай decision	выход	цель	abserr	callrate	indpro	lgovb	m2	ordano	ordbno	tblnus	termas	nondug	ecpi	ucpi	decpi	riskp	testru	fus	retlag	levelag	retlag	levelag
110	100.00	2	2	0	0.5	-0.1	12.9	10.5	0.4	2.4	0.5	0.2	100.0	1.0	3.6	0.5	-0.7	0.2	-0.8	-0.2	-0.6	-0.2
111	69.39	2	2	0	14.0	-11.5	98.6	100.0	5.1	60.4	25.2	6.8	84.1	32.1	67.1	38.8	-49.2	10.0	-59.0	-16.0	-51.5	16.0
112	95.54	1	1	0	-0.5	-0.5	-11.1	19.4	4.7	-7.8	-11.1	1.3	100.0	0.7	-2.5	-0.6	-6.8	-0.1	9.1	24.8	38.8	24.8
113	1.18	1	3	2	-7.0	44.4	-74.3	-4.3	1.4	-44.1	-30.6	15.5	-4.4	-13.1	-27.8	1.8	3.9	-5.9	84.9	1.9	100.0	1.9
114	40.04	1	2	1	-1.0	5.1	-33.6	23.5	4.3	-26.3	-3.9	2.6	1.6	-1.7	-9.1	-27.8	0.9	-2.0	27.9	37.6	100.0	37.9
115	87.75	1	2	1	-0.5	2.4	-12.4	23.1	3.7	-9.5	6.7	1.2	18.7	-1.0	-4.2	6.3	2.1	-1.3	12.3	11.4	100.0	11.4
116	57.76	2	2	0	0.3	4.4	40.0	13.2	2.4	25.5	-17.6	-19.3	100.0	-0.4	6.9	-37.1	-5.3	1.4	-13.5	-14.0	-49.9	-14.0
117	3.25	1	1	0	0.2	13.5	-30.2	0.8	-0.5	-12.6	4.4	5.8	-43.7	-3.5	-6.2	-18.2	-0.3	4.2	28.1	2.7	100.0	2.7
118	100.00	1	1	0	-0.1	-3.8	-2.3	100.0	0.6	5.0	-0.5	3.1	48.5	8.5	2.8	-1.8	0.1	24.5	-2.6	0.3	25.4	0.3
119	100.00	1	2	1	0.5	-2.4	-6.3	30.6	0.6	-6.0	30.4	12.4	58.8	0.9	-1.3	-10.2	-2.6	0.2	7.1	-3.1	100.0	-3.1
120	100.00	1	1	0	0.4	5.2	-0.5	33.1	0.8	0.3	8.1	0.1	0.7	2.5	0.3	0.4	0.5	11.5	10.6	0.7	100.0	0.7
121	100.00	1	1	0	-0.1	-4.0	-1.8	93.5	2.5	-0.3	38.7	0.6	9.6	6.6	0.3	2.6	-1.3	33.7	4.2	-5.2	100.0	-5.2
122	46.74	1	1	0	3.6	8.6	5.6	29.6	3.5	-18.2	37.4	-12.4	-48.7	-1.0	-3.9	9.4	-0.2	-2.8	66.0	2.6	100.0	2.6
123	99.60	1	1	0	1.0	0.2	16.3	39.3	0.5	-6.3	56.8	-5.4	-11.1	1.0	-1.1	7.0	-3.1	0.1	24.0	-4.5	100.0	-4.5
124	100.00	1	1	0	-0.4	-0.7	2.0	100.0	2.9	3.3	-1.1	-0.3	1.2	6.4	1.0	-1.2	0.3	1.4	-2.3	0.0	-2.4	0.0
125	100.00	1	3	2	0.2	1.0	9.3	90.2	-0.6	-0.6	15.3	0.5	-2.1	3.7	0.1	2.0	3.6	0.9	34.9	-2.5	100.0	-2.5
126	100.00	1	2	1	0.2	-0.8	9.2	29.2	0.3	-2.3	18.5	-3.7	5.6	0.6	-0.5	-4.1	-3.0	0.0	6.8	19.7	100.0	19.7
127	100.00	1	2	1	-0.4	-3.2	3.8	100.0	0.4	2.8	2.9	-2.0	0.1	10.1	1.4	-1.0	-0.1	0.3	-1.9	2.2	38.3	2.2
128	100.00	1	1	0	0.0	0.1	4.4	53.5	0.3	-0.6	25.8	-1.8	-1.7	3.9	0.1	1.9	0.4	0.3	20.0	3.0	100.0	3.0
129	100.00	1	1	0	-0.8	-5.2	5.2	100.0	-1.1	5.1	13.5	0.2	5.4	14.9	1.9	-1.7	0.1	4.1	2.7	-0.2	66.7	-0.2
130	50.25	1	0	1	-7.2	-93.6	26.6	84.4	5.1	26.5	-44.7	1.0	-65.0	54.1	15.3	-25.9	3.2	4.4	-83.3	-10.1	-100.0	-10.1
131	100.00	1	1	0	0.0	-2.4	10.4	100.0	3.0	-0.3	74.6	-1.8	11.3	6.5	0.3	10.8	-0.4	-0.3	5.6	-4.4	74.9	-44.4
132	100.00	1	1	0	0.0	-1.2	5.4	60.7	0.2	-0.4	100.0	0.2	0.2	3.4	0.0	3.8	0.8	0.6	4.3	-1.9	42.1	-1.9
133	100.00	1	1	0	-1.9	-7.0	4.9	100.0	0.8	6.9	2.3	-1.8	-1.3	27.3	1.3	-3.0	0.1	35.3	-6.6	-1.4	-6.2	-1.4
134	100.00	1	1	0	-2.3	16.8	9.1	100.0	23.9	7.8	53.6	-1.5	24.6	30.7	2.9	-3.4	0.1	17.2	83.1	-2.6	50.8	-2.6

Таблица 6.9. Анализ вклада переменных в решающую способность (KNET) для периода времени с января 1989 г. по январь 1991 г.

Данные по макроэкономическим факторам, использованные Ченом, Роллом и Россом, после предварительной обработки оказываются связанными с целевой переменной (доходом на NYSE) почти идеально линейными связями, на что указывают большой коэффициент смешанной корреляции при регрессионном анализе и очень большое значение отношения R . Выбранная архитектура сети с непосредственными связями между входами и выходами представляется для такой ситуации довольно удачной.

Однако, даже в «безрадостной» ситуации MBPN-сеть может превосходить метод OLS-регрессии в смысле показателя RMSE и коэффициента корреляции Пирсона. Более того, 6-3-1 сеть даже на новых данных дает более точный прогноз, чем оценка регрессии на уже ранее обработанных данных.

Очевидна высокая степень согласованности результатов, касающихся вклада отдельных переменных, которые дают обычная регрессия и многослойная сеть. Такое соответствие повышает нашу уверенность в правильности результатов и одновременно говорит о том, что, по крайней мере, линейная составляющая связи между доходом по индексу и выбранным фактором улавливается нейронной сетью вполне успешно.

Остается открытым вопрос о том, в какой степени проделанные исследования поддаются обобщению. Результаты выглядят обнадеживающе. Перспектива завоевать рынок с помощью нейронных сетей, конечно же, весьма привлекательна. С точки зрения академического исследователя еще более интересной представляется возможность найти «истинную» модель для дохода по акциям. Методы ARIMA, VAR, TAR, ARCH и им аналогичные хороши для получения точных оценок временных рядов, но мало что дают для концептуального понимания исследуемого явления.

С учетом всего сказанного нейронные сети уже не выглядят как черные ящики, как это обычно пытаются представить. Мы рассказали о двух перспективных эвристических подходах к оценке динамики функциональных связей между доходами на рынке акций и переменными, описывающими состояние рынка. Один из возможных способов определения этих зависимостей состоит в том, чтобы кластеризовать их с помощью однородного или нелинейного анализа главных компонент. На полученные в результате этого кластеры можно смотреть как на возможные сценарии макроэкономического поведения. Таким образом будет подготовлена почва для применения простых правил торговли, включающих зависимость от времени.

Управление международным портфелем

ИНТЕРНАЦИОНАЛИЗАЦИЯ ПОРТФЕЛЬНЫХ ИНВЕСТИЦИЙ

В последние годы на международных финансовых рынках произошли большие перемены. После отмены контроля за обменом валют и снятия многочисленных качественных и количественных ограничений на инвестиционные возможности институциональных и частных фондов возникла целая индустрия заграничных инвестиций.

Управляющий пенсионного или страхового фонда имеет теперь все больше свободы для инвестирования и на основных международных рынках (США, Лондон, Токио), и на новых интенсивно растущих рынках Юго-Восточной Азии, Латинской Америки и других.

С помощью инвестиций за границу инвестиционные менеджеры стремятся улучшить соотношения риска-дохода, характерные для отечественного рынка, и через диверсификацию портфеля получать более высокий доход при том же или даже меньшем риске.

Имея в виду эту цель, менеджер портфеля бывает заинтересован в формировании инвестиционного пула с тем, чтобы иметь эффективную межнациональную позицию. Это означает, что нужно учитывать возросшее влияние на риск и доход всего совокупного портфеля. Первые математические постановки этой проблемы были даны Марковицем и Линтнером (см. [188], [182]).

Иллюстрируя эти основные принципы, Леви и Сарнат [178] показали, как формируется эффективный международный портфель, который, вообще говоря, должен быть устроен иначе, чем обычные инвестиции. Так, в частности, развившаяся в послевоенные годы в США тенденция направлять инвестиции, главным образом, в Канаду и Великобританию, при всех удобствах в отношении языка, системы законодательства и устройства рынка, не является оптимальной, поскольку доходы, получаемые в Канаде, очень сильно коррелированы с доходами в США, а Великобритания в этом отношении стоит на втором месте. Таким образом, смешанный американско-канадский портфель инвестиций будет нести в себе почти такой же риск, что и чисто американский, но доход по нему будет значительно ниже, так

как прибыли на канадском рынке акций ниже, чем в США. Портфель, состоящий частью из акций американских компаний, а частью из малорисковых американских же облигаций, будет по всем показателям превосходить американско-канадский портфель акций.

Иначе обстоит дело на фондовом рынке Индии. Хотя из-за особенностей законодательства, банковской структуры и политики в области занятости доходы здесь ниже, чем в США, а риск выше, при этом, как показали Леви и Сарнат, доходы, как правило, мало коррелированы с положением на американском рынке. Портфель, состоящий из индийских и американских акций, будет, по сравнению с чисто американским портфелем, давать более низкий доход, но зато с ним связан гораздо меньший риск. В табл. 7.1 приведены показатели дохода и риска для различных стратегий диверсификации.

	<i>Средняя норма прибыли</i>	<i>Стандартное отклонение</i>
Развивающиеся страны	5.0	26.5
Общий рынок	15.0	25.0
Западная Европа	15.5	23.5
Страны с высокими доходами населения	13.0	12.5
Все государства	12.0	8.0

(данные взяты из [178, с. 673])

Таблица 7.1. Средние нормы прибыли и стандартные отклонения для оптимальных портфелей (в процентах) при 5-процентной ставке

В этих рассуждениях существенным является предположение, что мы рассматриваем относительно небольшие по объему инвестиции. Капитализация развивающихся рынков столь слабо связана с рынками США или Токио, что управлять оптимально диверсифицированным портфелем было бы практически невозможно. У транснациональной оптимизации есть, однако, одна любопытная особенность. По мере того, как фонды начинают перетекать с основных рынков на некоррелированные с ними рынки, денежная масса поднимает индекс на этих малых развивающихся рынках, т.е. увеличиваются доходы на них, и это еще более усиливает тенденцию перетекания капитала (см. [268]). Даже после того, как доходы на обоих рынках становятся более скоррелированными, полученный в результате притока капитала повышенный уровень доходов продолжает усиливать вес развивающихся рынков в интернациональном портфеле.

Процесс глобальной диверсификации породил возникновение таких новых рынков, как европейские рынки второй волны (Стамбул),

рынки Латинской Америки, Азии, Африки и Ближнего Востока. Для молодых рынков характерны высокая волатильность и высокие прибыли, и, поскольку большая доля их волатильности не коррелирована с развитыми рынками, выход на такие рынки не добавляет большого риска к уже существующему портфелю. В табл. 7.2 приведены некоторые показатели, характеризующие относительный вес основных рынков и лидирующих развивающихся рынков.

		Доля мирового рынка (%)	Индекс MSCI
Европа		28.2	30.7
	Великобритания	11.2	12.2
	Германия	4.1	4.5
	Франция	3.7	4
	остальные	9.2	10
Сев. Америка		39.3	42.7
	США	36.8	40
	Канада	2.5	2.7
Дальний Восток		24.4	26.5
	Япония	20.3	22
	Гонконг	1.7	1.8
	Австралия	1.6	1.7
	остальные	0.9	1
Другие части света		8.1	0.1
	Тайвань	0.1	
	Мексика	0.1	

(данные взяты из International Finance Corporation, Quarterly Review of Stock Markets, quarter 2, 1992; Morgan Stanley Capital International index; Datastream)

Таблица 7.2. Капитализация мирового рынка по частям света

Преимущества, которые можно получить за счет диверсификации на новых рынках, хорошо понимают исследователи (см. [102], [248], [283]) и, разумеется, профессионалы инвестиционного менеджмента.

СПОСОБЫ ОЦЕНКИ РЕЗУЛЬТАТОВ

Задача оценки эффективности международных инвестиций ставит ряд интересных вопросов, которые влияют на выбор спецификации нейронной сети. Первый вопрос касается базовой валюты. Очевидно, что американский инвестор измеряет доходы и капитал в долларах и будет заинтересован в долларовой прибыли, так что на доходе его портфеля будет отражаться валютный риск, свойствен-

ный любым заграничным инвестициям. Инвестор из другой страны может на полном основании совершенно иначе оценивать риск и доход от того же самого портфеля. Рисковый доход в долларах от недолларового актива зависит от курса этой ценной бумаги в местной валюте и от обменного курса. Специалисты по техническому анализу, разработавшие хитрые теории про уровни сопротивления на ценах, выраженных круглыми числами, не учитывали то обстоятельство, что круглая сумма в долларах не будет круглой в йенах или немецких марках. Поэтому большое (и все возрастающее) количество сделок совершается в блаженном неведении относительно этого психологического барьера. При построении нейронно-сетевой модели в качестве входных переменных можно брать показатели как в долларах, так и в местной валюте.

При структурировании международного портфеля, наряду с индексами местных рынков, выраженными в местных валютах, хорошим средством является Международный индекс капитализации Моргана Стэнли (Morgan Stanley Capital International index, MSCI). Этот индекс выражает относительный вес мировых рынков и вычисляется, исходя из их долларовой капитализации.

Когда капитал инвестирован в иностранные активы, изменения индекса MSCI страны прямо отражают относительную прибыльность направленных в нее вложений.

Портфель, перегруженный вложениями в какой-то из рынков (т.е. когда доля инвестиций в него превышает долю капитализации этого рынка), будет опережать индекс MSCI в случае, когда доля данного рынка в мировом масштабе возрастает, и, соответственно, отставать, если рынок сокращается. При этом во втором случае портфель может продолжать расти в долларовой цене, но это будет происходить медленнее по сравнению с портфелем, распределенным в точном соответствии с капитализацией рынков (такой портфель называется нейтрально взвешенным или индексным). Все это можно проиллюстрировать на условном примере, см. табл. 7.3.

	Индекс MSCI (%)	Индекс MSCI (%)	Портфель № 1	Портфель № 1	Портфель № 2	Портфель № 2
	t_{-1}	t_0	t_{-1}	t_0	t_{-1}	t_0
Страна А	30	35	40	46.6	20	23.3
Страна В	25	25	40	40.0	30	30.0
Страна С	45	40	20	17.7	50	44.4
всего	100	100	100	104.3	100	97.7

Таблица 7.3. Пример эффективности портфелей

В момент t_1 портфель № 1 по сравнению с индексом MSCI перегружен акциями компаний государств А и В и недогружен акциями страны С. Пусть доход в стране А выше индексного, в стране В — равен индексному, а в С — ниже индексного (но разница в процентах меньше, чем у индексного и А). В результате в момент t_0 портфель № 1 выигрывает у индексного в стране А и несколько проигрывает в С, что вместе дает преимущество в 4.3% перед индексным портфелем. Портфель № 2 перегружен за счет акций В (цена которых держится на уровне) и С (которые отстают). В результате этот портфель проигрывает индексному 2.3%.

Пере- или недогруженность сама по себе еще ничего не решает, потому что повышенный доход может быть получен только за счет увеличения риска портфеля. Под риском мы понимаем прошлую (историческую) волатильность временных рядов доходов (в данном случае — относительных доходов). Очевидно, что при перегруженности активами более волатильного рынка риск возрастает, и даже при более высоких прибылях отношение риск/доход может быть хуже, чем у индекса. Поэтому при сравнении эффективности нужно либо стандартизировать риск, либо сравнивать между собой портфели с близкими уровнями риска. Такие сравнения, однако, трудно выполнить, потому что управляющие фондами могут хеджировать или спекулировать на межвалютном риске своих портфелей, и становится неясно, каков же действительный риск. В качестве практического решения можно, не оценивая границ риска для составляющих портфеля по исторической эффективности, просто подсчитать риск непосредственно по волатильности портфеля.

ФОРМИРОВАНИЕ ПОРТФЕЛЯ: ЭКСПЕРТНОЕ МНЕНИЕ

На протяжении ряда лет журнал "Economist" публикует рекомендации по формированию портфеля, принадлежащие менеджерам ведущих транснациональных инвестиционных фондов. Раз в три месяца публикуются рекомендации менеджеров десяти разных фондов, а также нейтральные веса по индексу MSCI на этот момент. Рекомендации охватывают все аспекты работы на мировых рынках, причем отдельно рассматривается распределение активов между облигациями и акциями.

Эти прогнозы трудно оценить и точно реализовать по следующим двум причинам. Во-первых, параметры разделения средств между акциями и облигациями не поддаются буквальному истолкованию, поскольку инвестиционные компании, о которых идет речь, предоставляют возможность осуществлять платежи в различных валютах. Например, йеновые облигации несут гораздо меньший риск

(или даже являются безрисковыми) для японского инвестора, считающего прибыль в йенах, чем для американского, считающего ее в долларах. Поэтому мы оставим в стороне вопрос о разделении на акции/облигации и сосредоточимся на более простой задаче выбора наилучшего портфеля акций.

Вторая трудность заключается в том, что география некоторых капиталовложений может охватывать несколько регионов, так что такому вложению соответствует не один показатель индекса, а взвешенная комбинация нескольких. В наших нейронно-сетевых моделях мирового рынка все прогнозы формулируются в терминах отдельных рынков. В тех случаях, когда формат данных не позволяет отслеживать эффективность действий реальных менеджеров на отдельном рынке, мы просто производили перераспределение портфеля.

Таким образом, мы приходим к следующей, более узкой задаче: сравнить результаты нейронно-сетевой модели с ежеквартальным распределением активов, предлагаемым в журнале "Economist", по пяти основным фондовым рынкам: США, Японии, Великобритании, Франции и Германии. Совокупный объем этих рынков составляет около 85% от всего объема мирового рынка, который учитывается в индексе MSCI.

Кроме этого, для нас представляют интерес возможности нейронных сетей при работе с более широким набором рынков, поэтому было сделано расширение модели, чтобы охватить еще 11 региональных рынков, включая Гонконг, Сингапур и ряд европейских рынков.

СПЕЦИФИКАЦИЯ МОДЕЛИ

Поскольку публикуемые прогнозы принадлежат специалистам из разных стран, представляется затруднительным ввести единую спецификацию для риска. Очевидно, риск в йенах — не то же самое, что долларовый риск. Кроме того, анализ данных показывает наличие сильного избирательного эффекта: инвесторы чаще всего отдают предпочтение своим отечественным рынкам. Очевидно, в этом сказывается желание уменьшить риск в расчете на свою валюту, однако здесь отражается еще и тот факт, что инвестиционные компании считают себя лучше информированными о перспективах собственного рынка и, чтобы использовать это преимущество, увеличивают свои вложения в него.

По этим причинам мы решили не брать показатели эффективности риска/дохода в качестве цели прогноза, а выбрали более простой путь: управление риском происходит путем ограничения отклонений портфеля, сформированного нейронной сетью, от индексных весов.

Исходя из индексов MSCI, нейронная модель предсказывает изменения каждого из пяти целевых индексов на месяц вперед. Затем происходит перераспределение портфеля таким образом, чтобы он имел большие веса на наиболее сильно растущих рынках.

Делается это так: подсчитываются предполагаемые доходы за пять месяцев вперед, и затем вес портфеля на рынке с наибольшим доходом берется на 40% больше, чем соответствующий индекс MSCI. Для следующего по порядку рынка увеличение берется 20%, для третьего вес остается на прежнем уровне, для четвертого и пятого — уменьшается на 20% и 40%, соответственно. Так как рынки различаются размером, портфель нужно затем подправить с сохранением пропорций таким образом, чтобы суммарный вес был равен 100.

Остается последняя проблема: делать прогноз в долларах или в собственной валюте? Нас интересовала долларовая эффективность, отражающаяся в индексе MSCI, но при этом нужно выбрать, что лучше: давать прогноз для каждого рынка в его валюте, а затем переводить в доллары, или непосредственно прогнозировать долларовый эквивалент каждого индекса. В конце концов, мы пришли к тому, что разработали нейронные модели и для прогнозирования долларового значения национальных индексов, и для индексов в собственной валюте, которые затем можно пересчитать в доллары, пользуясь одномесячными форвардными обменными курсами.

ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА

Были собраны ежемесячные данные (в двух вариантах: долларом и в местной валюте) о фондовых индексах на 16 региональных рынках, а также данные MSCI. С самого начала стало ясно, что данные содержат высокий уровень шума. Были сформированы множества скользящих средних значений, охватывающих два, три и четыре месячных промежутка.

В качестве целевой переменной бралось четырехмесячное скользящее среднее для каждого индекса, выраженное либо в местной валюте, либо в долларах. Для примера в табл. 7.4 приведен набор переменных, соответствующих австралийскому рынку.

В другом варианте модели долларового прогноза для Австралии целевая переменная и переменная 10 менялись местами, с внесением соответствующих поправок в вычисление временных задержек.

Затем были сформированы файлы входных данных следующего уровня путем вычисления процентных изменений соответствующих скользящих средних. В выходном слое содержалось прогнозируемое на следующий месяц для данного государства процентное изменение четырехмесячного скользящего среднего.

Целевая переменная Долларовый MSCI-индекс Австралии для четырех-
месячного скользящего среднего $t_{-3}:t_0$

Описательные переменные

1	Индекс Японии в долларах; t_{-1}
2	Индекс Японии в местной валюте; t_{-1}
3	Индекс Великобритании в долларах; t_{-1}
4	Индекс Великобритании в местной валюте; t_{-1}
5	Индекс США в долларах; t_{-1}
6	2-периодное скользящее среднее для Австралий- ского индекса; доллар; $t_{-1}:t_2$
7	2-периодное скользящее среднее для Австралий- ского индекса; местн. валюта; $t_{-1}:t_3$
8	3-периодное скользящее среднее для Австралий- ского индекса; доллар; $t_{-1}:t_3$
9	3-периодное скользящее среднее для Австралий- ского индекса; местн. валюта; $t_{-1}:t_3$
10	4-периодное скользящее среднее для Австралий- ского индекса; местн. валюта; $t_{-1}:t_4$

Таблица 7.4. Модель для Австралии. Целевая и описательные переменные

ОБУЧЕНИЕ

Использовалась классическая 10-3-1 сеть с прямой связью и алгоритмом обратного распространения ошибки. Сеть имела 10 входных узлов, 3 узла в скрытом слое и выходной слой, состоящий из одного узла. Применялся программный пакет NevProp с параметрами, приведенными в табл. 7.5.

В случаях, когда делались прогнозы для США, Японии и Великобритании, в качестве целевого индекса выбирался индекс MSCI для Германии, представляющей второй по величине рынок. Это делалось для того, чтобы не иметь дело с проблемой корреляции с данными о скользящих средних для той же страны.

Обучающее множество состояло из 116 месячных скользящих средних, и еще 23 было оставлено для тестового множества. Обучение модели включало 1000 эпох с контролем процесса сходимости через каждые 5 эпох, а в качестве критерия сходимости брался квадратный корень из среднеквадратичной ошибки (Root Mean Square Error, RMSE) модели. Начальные значения весов были поло-

жены равными единице, за исключением трех из них, которые для лучшей сходимости были взяты равными нулю. Затем вычислялись изменения четырехмесячного скользящего среднего, прогнозируемые сетью, и из сравнения с текущим уровнем индекса находилось предполагаемое его значение на месяц вперед.

1	Weight Range	0.1
2	Hyper Error	0
3	Sigmoid Prime Offset	0
4	Epsilon	0.1
5	Split Epsilon	1
6	Momentum	0.1
7	Weight Decay	0
8	Score Threshold	0.1
9	QPMaхFactor	1.75
10	QPMoдеSwitch Threshold	0
11	Ninputs	10
12	Nhidden	3
13	Noutputs	1
14	UnitType	1
15	Connectcalls	2
16	NTrainingPatterns	116
17	NTestingPatterns	23

Таблица 7.5. Параметры обучения модели MSCI-индекса (названия соответствуют принятым в пакете NevProp)

РЕЗУЛЬТАТЫ

В табл. 7.6 приведены среднеквадратичные ошибки для 32 моделей.

Несколько неожиданным оказалось то, что ошибка прогнозов в местной валюте (т.е. без учета эффектов обмена валют) мало отличалась от ошибки долларовых прогнозов. В отдельных случаях, в частности, для Канады и Австралии, ошибка для местной валюты была выше, чем для долларового прогноза. В применении к интернациональному портфелю обменным риском можно управлять, приняв предположение о равенстве покупательных способностей и применяя к совокупному портфелю модель CAPM с одним бета (см. [4]), или же, не предполагая равенства покупательных способностей, используя для каждой валюты свой показатель риска. Модели с несколькими факторами риска были предложены Россом и Уолшем [232].

Имея в виду наши цели, нет веских оснований ожидать, что пересчет местных цен в доллары по прогнозируемым обменным курсам

даст значительное улучшение точности, так как для этого нужен почти идеальный прогноз обменного курса на месяц вперед. Поэтому в качестве основы мы брали долларовый прогноз.

	Кв. корень из среднеквадратичной ошибки (%)	
	долларовое значение индекса	индекс в местной валюте
1 Япония	2.06	20.1
2 Великобритания	1.27	1.01
3 США	0.44	—
4 Франция	1.21	1.16
5 Дания	1.32	1.42
6 Канада	1.04	0.79
7 Бельгия	0.99	1.07
8 Австралия	1.68	1.32
9 Германия	1.33	1.40
10 Италия	2.12	2.27
11 Нидерланды	1.17	1.10
12 Норвегия	1.67	1.79
13 Сингапур/Малайзия	1.36	1.32
14 Швеция	1.82	2.24
15 Швейцария	1.28	1.21
16 Гонконг	2.88	2.83

Таблица 7.6. RMSE для 32 моделей

АНАЛИЗ РЕЗУЛЬТАТОВ

В качестве проверочных данных использовались данные с января 1992 г. по конец ноября 1993 г. Результаты тестов приведены в табл. 7.7.

Следующим шагом было сравнение результатов, которые выдавала сеть, с опубликованными «эталонными» портфелями. В качестве последних мы взяли ежеквартальные публикации журнала "Economist", общим числом около 80, для 8 периодов времени.

За два года степень детализации публикуемых портфелей существенно уменьшилась, многие мелкие рынки исчезли из рассмотрения или попали в категорию «прочие», так что их стало невозможно непосредственно связать с нашими временными рядами и прогнозами, относящимися к отдельным государствам.

При этом, однако, для всего периода времени имелись данные о рекомендованных весах вложений в рынки США, Японии, Великобритании, Германии и Франции. Именно эти данные мы использовали для сравнения, результаты которого приведены в табл. 7.8.

В первом числовом столбце табл. 7.7 указано количество портфелей, предложенных инвестиционными банками. Остались неучтенными рекомендации еще 6 специалистов (по одной у каждого).

	Кол-во портфелей	Средний анормальный доход	Кв. корень из среднеквадр. ошибки (в %)	Абсолютное отклонение от ней- тральных весов (в %)
Индекс MSCI	8	-0.50	(0.43)	0.0
Daiwa Europe	8	-0.44	(1.28)	32.3
Merill Lynch	7	-0.32	(0.89)	13.3
Lehman Brothers	7	+0.09	(2.03)	29.9
Credit Agricole	7	-0.18	(4.74)	43.0
Nikko Securities	7	-0.20	(0.69)	11.6
Credit Suisse	6	-0.28	(1.28)	30.6
Commerz International	8	-0.35	(2.39)	33.5
UBS/Philips and Drew	7	-0.30	(2.40)	56.5
Robeco Group Asset Management	7	-0.27	(1.30)	32.6
Julius Baer Zurich	7	+0.31	(6.65)	52.1
Нейронная сеть	8	+0.10	(0.54)	14.6

Таблица 7.7. Сводка результатов

	Доход	Риск	Отношение риск/доход
Индекс MSCI	0	0.43	—
Daiwa Europe	0.06	1.28	14.90
Merill Lynch	0.18	0.89	4.90
Lehman Brothers	0.59	2.03	3.40
Credit Agricole	0.38	4.74	12.50
Nikko Securities	0.30	0.69	2.30
Credit Suisse	0.22	1.28	5.80
Commerz International	0.15	2.39	15.90
UBS/Philips and Drew	0.20	2.40	12.00
Robeco Group Asset Management	0.23	1.30	5.65
Julius Baer Zurich	0.81	6.65	8.21
Нейронная сеть	0.60	0.54	0.90

Таблица 7.8. Сравнение риска и дохода по отношению к индексу MSCI

В качестве точки отсчета был выбран индекс MSCI. Мы рассматривали инвестиции на трехмесячный срок без непрерывной корректировки портфеля в соответствии с рынком. Поэтому веса, которые были нейтральными относительно MSCI в начале трехмесячного срока, уже не будут таковыми к его концу — на анализируемом отрезке времени трехмесячная задержка давала по сравнению с непрерывно индексируемым портфелем проигрыш примерно в 0.5%.

Все профессиональные менеджеры превзошли MSCI-нейтральный портфель, проиндексированный в начале трехмесячного периода, однако все, за исключением Lehman Brothers и Julius Baer, дали отрицательный анормальный доход в сравнении с полной индексацией. Особую важность имеет то обстоятельство, что большинство рекомендованных портфелей имеет высокую среднеквадратичную ошибку (RMSE). У всех, кроме Nikko Securities, она намного выше, чем у взятого за точку отсчета индекса MSCI.

Нейронная сеть формировала свой портфель, исходя из прогнозируемых ею показателей дохода на месяц вперед по пяти рынкам. Результат оказался хороший — доход немного выше, чем у Lehman Brothers, а более высокий показатель имеет только портфель Julius Baer.

Видно, однако, что нейронная сеть достигает таких результатов при гораздо меньшей среднеквадратичной ошибке, чем все авторы рекомендаций. Ее RMSE лишь немного больше, чем у трехмесячного индекса MSCI, а, например, у Julius Baer и Lehman Brothers риск получился намного больше.

Причину этого отчасти объясняют данные последнего столбца табл. 7.7. В среднем нейронная сеть дает меньшее среднее отклонение от индексного портфеля, чем все другие портфели, кроме Merrill Lynch и Nikko Securities. 43-процентное отклонение портфеля Credit Agricole от нейтрального означает, что распределение активов этого банка в среднем в каждой стране отличается на 8% от индекса MSCI. Большая часть этого отклонения приходится на четырехкратно перегруженную позицию по французским и немецким акциям за счет сокращения доли акций Японии и США.

ВЫВОДЫ

Нейронная сеть вполне успешно улавливает и внутреннюю динамику, и корреляцию между национальными и основными глобальными рынками. Ввиду коротких сроков прогноза волатильность обменных курсов, по-видимому, существенно не сказывается на точности прогнозов: RMSE примерно одинакова для долларового варианта и для местной валюты. Это подтверждает оправданность использования сетью обоих временных рядов.

Среди всех рынков наибольшие трудности у модели возникли с Гонконгом, Швецией, Италией и Японией, тогда как сделать прогноз для США на этом отрезке времени оказалось особенно легко.

В сравнении с тем, как разместили свои активы профессиональные менеджеры, результаты нейронной сети выглядят обнадеживающе: сеть добилась наивысшего дохода при наименьшем (не считая индекса) риске среди всех портфелей. Фактически сеть добилась аномального дохода на 0.6% в квартал (около 2.5% в год) выше, чем дал бы ежеквартально индексируемый портфель, и примерно на 0.5% в квартал выше непрерывно индексируемого. При этом мы использовали только треть выдаваемых сетью прогнозов, а если бы мы на основании месячных прогнозов производили ежемесячную корректировку портфеля, то результаты были бы еще лучше. Представляется, что на данном этапе сеть дает вполне доброкачественный прогноз.

8

Оценка кредитного риска на основании данных нефинансового характера

Многие модели предсказания банкротств убедительно демонстрируют свои способности к прогнозированию в задачах классификации задним числом (*ex post*). В ситуации же *ex ante*, когда права при банкротстве, резервы выживания и кредитные резервы взаимосвязаны, эти модели работают плохо. С другой стороны, как заметил Арженти, при том, что нейронные сети могут работать и с числовыми, и с нечисловыми данными, было предпринято очень мало попыток включить в рассмотрение данные качественного характера.

В описанном далее исследовании мы ставили перед собой цель построить нейронно-сетевые классификаторы для двух задач. Первая задача состояла в оценке финансовых трудностей компаний в ситуации *ex ante* на основе информации, которой располагали 12 финансовых чиновников в Польше. В восьми ведущих польских банках была собрана база данных по 59 (анонимным) компаниям, включающая 21 качественный и 5 количественных показателей. Наличие двух типов данных не позволяет применить обычные дискриминантные модели типа Logit или MDA, поэтому данные были преобразованы с помощью нелинейного анализа главных компонент. Такое преобразование дало нам возможность сравнивать результаты нейронно-сетевой модели, полученные при перекрестном подтверждении (кросс-валидации), с результатами линейного метода MDA.

Сравнение (в смысле ошибок 1-го и 2-го рода) результатов *ex ante* — прогнозирования с помощью нейронной сети и на основе традиционных методов MDA — свидетельствует о том, что применение нейронных классификаторов оправданно в условиях постоянно меняющейся ситуации, что характерно для польского рынка капиталовложений.

Вторая прикладная задача, которая рассматривается в этой главе, связана с *ex post* оценкой кредитного риска по фрагменту портфеля займов, выданных корпоративным клиентам Голландского Инвестиционного банка (NIB). Этот частный торговый банк с преобладающей долей государственной собственности специализируется на предоставлении средне- и долгосрочных займов корпорациям. Для описания компаний-заемщиков банка в модели использовалось 44 каче-

ственных и 5 количественных переменных. Сведения о компаниях, которые обращались с просьбой о предоставлении займа и которым было отказано, не сохранялись, поэтому проведение *ex post* анализа ошибок 2-го рода было невозможно.

МОДЕЛИ ПРЕДСКАЗАНИЯ БАНКРОТСТВ

Грэдди и Спенсер [128] предложили выделить в кредитном риске 5 составляющих:

- 1 Желание клиента погасить долг,
- 2 Способность выплатить долг,
- 3 Чистый капитал компании,
- 4 Залог,
- 5 Чувствительность к изменениям положения в экономике.

Коль скоро кредитный риск определяется и качественными, и количественными факторами, в модели оценки риска следует учесть и те, и другие. При этом в отношении качественных показателей возникает вопрос об их объективности и о возможности их измерения. Из-за того, что обычно используемые в таких задачах статистические модели (MDA, логистическая регрессия) не приспособлены для работы с данными качественного характера, последние вынужденно выпадают из *формальной* процедуры метода. Таким образом, почти все модели анализа кредитов строятся на количественном подходе, хотя какие-то элементы качественного анализа также могут присутствовать. Модели, дающие «точное» решение, получаются довольно невразумительными, дилетантскими и не способными определить, в какой степени качественные составляющие действительно влияют на результат. Поэтому с их помощью очень трудно оценить решения, принимаемые специалистами в этой области.

При использовании количественных моделей исходят из того, что шансы компании на выживание можно оценить, отслеживая, как меняются со временем соотношения между различными финансовыми показателями. Обсудим вкратце два наиболее известных метода такого рода — множественный линейный дискриминантный анализ (MDA = Multiple Discriminant Analysis) и регрессию.

MDA — это статистический метод изучения различий между двумя или более группами объектов по совокупности нескольких финансовых показателей. Объекты (в данном случае компании, обратившиеся с просьбами о предоставлении займа) разбиваются на несколько попарно непересекающихся групп на основании ряда показателей, характеризующих надежность и успешность их работы. Основные предположения здесь состоят в том, что имеется два или более возможных исходов и что описывающие модель переменные

распределены в своем интервале изменения в соответствии с многомерным нормальным распределением. Прогнозирование банкротств с помощью MDA-моделей было популярно в Великобритании и США в 70-е годы (см. [12], [257]). Затем выяснилось, что наборы определяющих переменных у разных моделей сильно различаются, и MDA-методы отчасти потеряли свою привлекательность. Оптимальной процедуры выбора совокупности переменных не существует, однако некоторые исследователи с помощью факторного анализа уменьшают размерность задачи, оставляя только те переменные, которые сильно коррелируют с конечным результатом (см. [180], [11]). Барнс [29] связывает чрезмерное многообразие описательных переменных с тем фактом, что процедура отбора переменных теоретически не согласуется с моделями, основанными на статистических критериях согласия. Более того, остается открытым вопрос о том, можно ли пользоваться многомерным нормальным распределением применительно к финансовым показателям, которые в действительности распределены не нормально (ср. [155], [226]).

Другой подход к выявлению возможности банкротства основан на регрессионном анализе. В отличие от MDA, линейные вероятностные модели (LPM = Linear Probability Models) и логистическая регрессия (logit) предполагают причинную связь, идущую от экзогенно определяемых переменных и случайных ошибок к зависимой бинарной переменной. К сожалению, четких правил для оценки LPM-моделей нет. Особенно неприятно то, что линейно регрессионные варианты LPM-моделей могут выдавать отрицательные или превышающие единицу оценочные значения для вероятностей. Модели probit и logit (соответственно, со стандартной нормальной и логистической функцией преобразования) в этом смысле лучше, потому что преобразование является монотонным, его выходные значения ограничены нулем и единицей и стремятся к нулю и единице на хвостах распределения. Все это согласуется с точкой зрения, что никакое событие, даже если речь идет о далеких выбросах, не может быть предсказано с абсолютной уверенностью, т.е. с вероятностью нуль или единица. Гильберт, Менон и Шварц [125] приводят типичный пример logit-регрессии:

$$\Pr(Y) = 1/[1 + \exp\{-(\alpha + \beta X)w\}], \quad (1)$$

где $\Pr(Y)$ — вероятность банкротства; X — множество независимых переменных; α , β — коэффициенты, которые требуется оценить.

Так как Y принимает значения от 0 до 1, их обычно интерпретируют как вероятность принадлежности результата к данному классу. Сравнивая logit-методы с MDA, Коллинз и Грин [73] утверждают, что, хотя logit-методы дают меньшую ошибку 1-го рода, в целом точность классификации у них ненамного лучше. Учитывая все это,

мы выбрали метод MDA в качестве точки отсчета для оценки результатов, которые выдает сеть.

Арженти [15] первым отметил важность качественных показателей в вопросе о банкротстве корпораций и выделил 12 переменных, из которых 8 являются причинными факторами, а остальные 4 — симптомами банкротства. Наиболее важными причинными факторами являются плохое управление (авторитарный стиль) и некачественная система информации, тогда как ухудшение финансовых показателей и подтасовка отчетности — симптомы ухудшения положения. Джордж [121] предложил аналитическую схему, включающую два нефинансовых показателя: компетентность управления и стратегическое положение. Прогнозов автор не делал, но он предполагает, что качественный анализ факторов делового риска может дополнять количественный анализ. Кизи и Уотсон [158] проверили гипотезу Арженти на реальных данных. Среди 18 переменных их модели были как переменные типа ДА-НЕТ, так и непрерывно меняющиеся переменные (средний интервал между аудиторскими проверками в последние 3 года, число членов в совете директоров в настоящий момент и др.). На материале данных о двух группах предприятий (обанкротившихся и сохранившихся) одномерный анализ выявил значительные различия в значениях переменных внутри групп. Что особенно важно, качество классификации логистической регрессионной модели заметно улучшилось после включения в нее нефинансовых переменных Арженти.

Все это говорит о том, что такие сравнения нейронно-сетевых и MDA-моделей полезны, так как мы получаем возможность сопоставить новый, более устойчивый метод с известной моделью, в которой приняты априорные предположения о переменных. В этой связи отметим применения нейронных сетей к задачам выявления возможных нарушителей налогового законодательства (см. [163]), банкротств промышленных корпораций (см. [223]) и банкротств финансовых корпораций (см. [259]). Другие приложения можно найти у Триппи и Турбана [266]. При этом во всех случаях в качестве входных данных использовались только стандартные (полученные по методу Альтмана) числовые переменные. Поскольку эти финансовые показатели отражают, скорее, прошлое состояние, мы вместо этого предпочитаем использовать качественные переменные, которые, как нам представляется, несут в себе информацию о будущем.

ПРЕДОСТАВЛЕНИЕ ЗАЙМОВ МАЛЫМ И СРЕДНИМ ПРЕДПРИЯТИЯМ (ОПЫТ ПОЛЬШИ)

Малые и средние предприятия (МСП) Польши в настоящее время испытывают большие трудности с финансированием, вызванные

экономическим спадом, высокими процентными ставками, установленными ограничениями и балансовым дефицитом. Пытаясь выжить, МСП обратились за поддержкой к банковскому сектору. Предоставит ли финансовая организация свой капитал для размещения на данном МСП, зависит от его характеристик риска и от положения дел в секторе МСП в целом.

С точки зрения кредитора степень рисковости займа зависит от рисковости фирмы-заемщика и от условий соглашения о займе (срок, проценты, обеспечение). Оценка рисковости МСП, обратившегося с просьбой о займе, является для польских банков непростой задачей по двум причинам. Во-первых, при коммунистическом режиме средства, хотя и переводились со счета на счет, оставались при этом государственной собственностью, и поэтому кредитный анализ был бессмыслен. Теперешние быстрые изменения в системе коммерции заставили банкиров приобрести необходимую квалификацию в области оценки займов (например, была открыта Международная банковская школа в Катовицах). Во-вторых, непрерывно меняющиеся правила финансовой деятельности вынуждают специалистов, делающих прогнозы, принимать многочисленные и порой рискованные предположения о будущем состоянии дел на МСП. Наконец, даже прошлая финансовая информация может быть недоступной (например, для недавно возникших фирм) или недостоверной (из-за устаревших методик подсчета).

В итоге банки пришли к тому, чтобы не решать напрямую вопрос о степени рисковости фирмы, а добиваться нужного результата через условия займа. Как следствие, возможность получения займа оказалась неразрывно связанной с возможностью фирмы предоставить под него обеспечение (залог). Конкретные факты таковы.

- Общая продажная стоимость залога должна быть не менее, чем в 1.5 раза (обычно в 2.5 раза) больше суммы займа.
- В качестве обеспечения займа принимаются только основные средства: земля, товары, имущество. Могут быть также приняты гарантии третьей стороны (особенно, если эта сторона — другой банк) или достаточный по размерам депозитный вклад. Активы типа акций или долговых расписок в залог не принимаются никогда.
- Сроки займов, как правило, невелики, в среднем — один год.
- Молодые компании фактически не имеют доступа к капиталу, если только они не будут финансировать 35% (иногда доходит и до 50%!) суммы инвестиций из своих наличных ресурсов.

Нерасположенность к риску, характерная для польских банков при работе с МСП, косвенно влияет на инвестиционную политику

и иностранных инвесторов. Такие организации, как Всемирный банк, Европейский инвестиционный банк и Европейский банк реконструкции и развития предлагают своим польским партнерам кредитные линии в надежде на то, что средства будут вложены в сектор МСП, где они так необходимы. Однако из-за позиции местных банков, которые не склонны оперативно рассматривать заявки на предоставление займа и смотрят на МСП, скорее, как на источник неприятностей, а не прибыли, польские МСП едва ли могут рассчитывать на помощь иностранных инвесторов. Специально для того, чтобы попытаться заполнить эту брешь, была создана бесприбыльная организация Enterprise Credit Corporation — дочерняя фирма Польско-американского фонда предпринимательства (РАЕФ).

Недостаток капиталовложений не создавал бы такой проблемы, если бы доля сектора МСП в польской экономике была малой. К сожалению, в отличие от таких бывших коммунистических стран, как Венгрия и Чехословакия, в экономике Польши на частный сектор приходится значительная часть. Доля МСП в экономике с 1991 г. по 1992 г. возросла с 26% до 31%. Всего за год число занятых в этом секторе увеличилось с 24.1% до 58.5% (в 1992 г.). Даже если исключить сельское хозяйство, эта доля останется очень высокой — 44% (см. [95], [96], [166]). Поэтому необходимо изменить неблагоприятное для МСП положение в области кредитования и гарантировать им доступ к инвестициям. В идеале процедура оценки должна учитывать особенности данного сектора экономики, должен быть устранен упор на предоставление залога, увеличены сроки кредитов и приняты меры для поощрения молодых предприятий. Чисто количественные методы при решении вопроса о выделении кредитов здесь не вполне подходят из-за отсутствия прошлых данных и нестандартных способах учета в имеющихся базах данных. Один из возможных подходов к работе с качественными и отсутствующими данными — нейронные сети.

ОПИСАНИЕ БАЗЫ ДАННЫХ

С целью выделить факторы, определяющие кредитный риск для МСП (фирм с числом работающих менее 50), мы рассмотрели 23 польских финансовых организации (их полный список приведен в приложении 1 к данной главе). Это исследование, являющееся продолжением сделанного в работе Джорджа [121], дало основу для построения содержательной базы данных, включающей 60 гипотетических наблюдений для 26 описывающих показателей. Затем были приглашены специалисты в области кредитного дела из 14 новых финансовых организаций, и им было предложено оценить привлекательность инвестиций в МСП (под привлекательностью понимается

способность фирмы вернуть долг с процентами), ответив на следующие три вопроса:

- Считаете ли вы, что это МСП столкнется с финансовыми трудностями в течение ближайших 6 месяцев?
- Считаете ли вы, что это МСП столкнется с финансовыми трудностями в течение ближайших 18 месяцев?
- Изменили бы вы свою оценку, если бы в деле присутствовал полноценный залог?

С нами согласились сотрудничать 12 специалистов (43%). Мы рассчитывали на их способности к «*ex ante*» прогнозированию финансовых трудностей. В качестве зависимой переменной выбирался усредненный прогноз того, будет ли данная фирма испытывать финансовые трудности в ближайшие 6 месяцев. Так как мы не располагали никакой информацией о вероятности принадлежности фирм к группам, то мы изначально взяли вероятности одинаковыми для всех групп. По результатам усредненного прогноза компании были разбиты на 3 группы:

группа 0	40 случаев	фирмы, которые должны просуществовать 6 месяцев
группа 1	16 случаев	пограничные случаи, разноречивых во мнениях экспертов
группа 2	3 случая	фирмы, которые должны обанкротиться в течение 6 месяцев

Один из случаев был отброшен, потому что целевую переменную невозможно было вычислить. В целом все эксперты оценили базу данных как содержательную и прислали свои ответы по факсу в течение 10 дней. Некоторые из них отметили сложность решения вопроса о кредитоспособности вне связи с его обеспечением.

В табл. 8.1 представлено репрезентативное подмножество гипотетической базы данных, которое было предложено экспертам. Совокупность переменных включает 21 нефинансовый показатель, значения которых обозначаются буквосочетаниями, и 5 финансовых показателей.

Все 26 переменных разбиваются на 4 группы:

- A Признаки данной компании или отрасли
- B Организационные факторы
- C Макроэкономические показатели
- D Финансовые показатели

	Компания	1	2	3
Отрасль промышленности	A1	TD	CO	MA
Рыночный потенциал	A2	5	5	4
Положение на рынке	A3	3	2	2
Экспорт	A4	N	N	Y
Помещения: собственность/аренда	A5	RO	RO	R
Возраст компании (лет)	A6	2	9	7
Структура собственности	B1	LL	LL	LL
Связь управляющий/владелец	B2	5	2	5
Централизованное принятие решений	B3	5	1	3
Гибкость организации	B4	4		1
Число работников (в т.ч. менеджеров)	B5	5	40	18
Возраст управляющего	B6	41	38	28
Стаж его работы в бизнесе (лет)	B7	10	7	4
Стаж работы в этой фирме (лет)	B8	15	6	6
Благонадежность управляющего	B9	4	1	2
Зависимость от обменных курсов	C1	2	2	5
Зависимость от инфляции	C2	1	2	5
Зависимость от таможенных сборов	C3	3	1	5
Зависимость от прямых налогов	C4	5	5	4
Зависимость от косвенных налогов	C5		2	2
Экологическая безопасность	C6	3	5	4
Чистая маржа	D1		-0.1	0.1
Коэффициент покрытия долга	D2	1.4	1.0	
Отношение ликвидности	D3	1.2	0.8	0.5
Рост объема продаж	D4	-0.3	-0.9	-0.1
Дебиторский показатель	D5	18	26	27
ПРОГНОЗ КРЕДИТОСПОСОБНОСТИ НА 6 МЕС.				

Таблица 8.1. Данные по трем гипотетическим МСП, предложенные экспертам

ГРУППА А: Признаки данной компании или отрасли (6 переменных)

Отрасль промышленности (обозначается двубуквенным сокращением) является очень важной переменной, поскольку от нее непосредственно зависят значения финансовых показателей и чувствительность фирмы к изменениям окружающей обстановки. Мы не брали межотраслевые компании, чтобы не было возможной потери информации. Были выделены 8 отраслей: торговля (TD), услуги, пищевая промышленность, легкая промышленность (МА), сельское хозяйство, строительство (СО), транспорт и прочее. Предполагалось, что компании не очень отличаются друг от друга размерами. *Рыночный потенциал* учитывает спрос на товары или услуги, производи-

мые фирмой, и зависит от того, в какой области работает компания. *Положение на рынке* оценивает соотношение цена/качество и конкурентноспособность фирмы. *Экспорт* (обозначение буквенное: Y/N) учитывает степень участия фирмы в экспортно-импортных операциях. *Помещения: собственность/аренда* (обозначение буквенное: O/R) указывает, владеет ли компания основными производственными помещениями. С арендой в Польше очень трудно из-за крайне высоких ставок, очень коротких (в пределах года) сроков и возможности отдавать недвижимость в залог. *Возраст компании* делит фирмы на старые (> 3 лет) и молодые.

ГРУППА В: Организационные факторы (9 переменных)

Структура собственности (буквенное обозначение) указывает на одну из семи категорий: совместное предприятие, предприятие с ограниченной ответственностью (LL = Limited Liability), гражданское товарищество, зарегистрированное товарищество, товарищество с ограниченной ответственностью, частное предприятие и частное лицо, зарегистрированное как фирма. *Связь управляющий/владелец* характеризует, насколько эти два лица связаны (одно и то же лицо, родственники, друзья, знакомые, не связаны). *Централизованное принятие решений* и *гибкость организации* в динамичных экономических условиях — очень важные факторы, выделенные Арженти [15]. *Благонадежность управляющего* показывает, насколько эксперт уверен в том, что управляющий будет считать себя обязанным вернуть долг с процентами. При решении этого вопроса польские банки без колебаний полагаются на сведения, полученные из полиции. Смысл остальных четырех переменных ясен без комментариев: *возраст управляющего*, *число работников*, *стаж работы управляющего на руководящей работе* (a) в бизнесе вообще, (b) — на данной фирме.

ГРУППА С: Макроэкономические показатели (6 переменных)

Шансы МСП на выживание зависят от следующих важнейших макроэкономических факторов: *инфляция*, *прямые налоги*, *косвенные налоги*, *таможенные и акцизные сборы*, *курсы обмена иностранных валют*. Несомненно, что введение в июле 1993 г. налога на добавленную стоимость поднимет уровень цен, и такое изменение будет очень чувствительным в ситуации, когда покупательная способность и рентабельность падают. *Экологическая безопасность производства* в настоящий момент, может быть, не имеет решающего значения, однако ввиду того, что южные районы Польши сильно загрязнены,

экологическая безопасность проекта может положительно повлиять на решение о предоставлении кредита. Для поддержки «зеленых» проектов была создана специальная общественная организация — Национальный фонд защиты окружающей среды и водных ресурсов.

ГРУППА D: Финансовые показатели (5 переменных)

Финансовые показатели заимствованы из «Z-модели» MDA Альтмана. *Чистая маржа* — это отношение чистого дохода к обороту. *Коэффициент покрытия при обслуживании долга* — это отношение прибыли до уплаты процентов и налогов (EBIT) к затратам на выплату процентов. Отношение ликвидности фирмы к сумме долговых обязательств выражает *степень ликвидности* позиции фирмы. *Рост объема продаж* — относительное (процентное) увеличение общего объема продаж — можно рассматривать как трендовую переменную. Наконец, *дебиторский показатель* — это средний срок платежей в днях (средний дебет / [чистый объем продаж × 360]).

Для обработки данных использовалась MBPN-сеть с логистическими функциями активации. Предполагалось, что после обучения сеть будет в состоянии правильно классифицировать новые (незнакомые ей) компании. В качестве исходной точки для сравнений была взята обычная линейная MDA-модель. Однако для метода MDA требуется, чтобы переменные были числовыми, — с буквенными или порядковыми переменными он работать не может. Проблема сведения всех показателей к числовым была решена при помощи нелинейного анализа главных компонент.

Нелинейный анализ главных компонент

С помощью нелинейного анализа главных компонент (АГК) мы не только преобразовали буквенные и порядковые переменные в числовые, но и уменьшили размерность множества данных с 26 (число переменных) до 5 (число значимых факторов). После этого, конечно, становится труднее представить себе суть этих новых составных переменных и понять, какое влияние каждая из них оказывает на результаты классификации. При АГК для каждого наблюдения вычисляются определенные числовые показатели этого объекта в каждом значимом измерении. Эти показатели (которые можно назвать количественными выражениями того, обладает ли объект тем или иным свойством) и используются в качестве входных данных для MDA. В итоге АГК дает новый набор данных меньшей размерности, чем у исходного (5 вместо 26), где уже все переменные являются числовыми. Конечно, эти два набора данных тесно связаны, поскольку

пять новых измерений охватывают большую часть степеней свободы, имевшихся в исходных данных. В отличие от MDA, которая не может работать с данными в исходном виде, сеть способна воспринимать информацию в различной форме. Поэтому мы брали не одну, а две сети и обучали одну из них на исходном множестве данных, а другую — на преобразованном. По результатам второго эксперимента можно сравнивать между собой качество работы сети и MDA.

АГК был выполнен с помощью модуля обработки категорий пакета SPSS+/PC, версия 5.01. В приложении 2 к этой главе приведены веса компонент в 5 значимых измерениях и собственные числа. Для примера там были взяты 10 первых компаний.

MDA КАК ТОЧКА ОТСЧЕТА

Используя полученные из АГК числовые характеристики объектов, мы провели стандартный линейный множественный дискриминантный анализ с одинаковыми (равными 33%) априорными вероятностями принадлежности элемента группам. Правильно были классифицированы 41% от общего числа случаев, и это несколько лучше 33-процентной точности, которая получилась бы при случайном отнесении объекта к той или иной группе. Табл. 8.6 ниже — это таблица неправильных классификаций, которая также называется матрицей ошибок.

Объект	Размерность				
	1	2	3	4	5
1	0.79	0.91	1.18	-0.68	-1.25
2	1.12	1.04	1.15	0.91	0.47
3	-1.07	0.53	0.07	1.08	-0.29
4	-0.94	-0.25	-0.31	-1.1	0.95
5	0.95	0.25	1.89	-0.4	0.39
6	-0.93	-0.38	0.03	-0.05	0.38
7	1.26	-1.63	-0.46	0.66	2.02
8	1.26	0.11	-1.77	-1.07	-0.99
9	-0.28	0.98	-1.87	1.78	-2.02
10	1.02	-2.11	-0.4	-0.54	-1.39

Таблица 8.2. Числовые характеристики объектов

Показатель ошибки 41% нуждается в уточнении. Дело в том, что ошибки бывают двух типов. Ошибка 1-го рода — это когда заем предоставлен и фирма обанкротилась, а 2-го рода — когда в предоставлении займа было отказано, и напрасно, потому что он принес бы

прибыль. Для банка, рассматривающего вопрос о предоставлении в первый раз займа новому клиенту, особенно важно минимизировать ошибки 1-го рода, так как они, очевидно, обходятся более дорого. В целом же проблема сложнее, поскольку отказ в предоставлении кредита постоянному клиенту может привести к тому, что он будет для банка потерян, и даже вызвать банкротство компании, которая смогла бы выжить, имея этот кредит. Таким образом, при минимизации ошибок 1-го рода возрастают ошибки 2-го рода.

НЕЙРОННО-СЕТЕВЫЕ МОДЕЛИ

На данной задаче были опробованы различные нейронные классификаторы, и результаты сравнивались с тем, что дают классические дискриминантные методы. Мы рассматривали многослойные сети (MBPN = Multilayer Backpropagation Networks), причем ограничились трехслойными сетями с одним скрытым слоем. Коэффициент обучения был взят равным 0.1.

Результаты для случая 5-мерной входной матрицы

Для значений целевой переменной (усредненного экспертного прогноза финансовых трудностей в ближайшие полгода) была выбрана следующая числовая запись:

группа 0	выживут	предоставить	0
группа 1	пограничные случаи	не знаю	0.5
группа 2	банкроты	отказать	1

Таким образом количественно выражается естественное отношение упорядоченности между тремя классами. Пять количественных показателей никаким преобразованиям не подвергались, поскольку их значения располагаются приблизительно от -2 до 2 и вполне могут быть поданы на вход сети в исходном виде. И у элементов скрытого слоя, и у выходного элемента функция преобразования была взята сигмоидной с единичной крутизной. Выходные значения разбивались на три категории, соответствующие предполагаемому ответу на основной вопрос:

$0 \leq \text{выход} \leq 0.33$	положить выход равным 0
$0.33 \leq \text{выход} \leq 0.66$	положить выход равным 0.5
$0.66 \leq \text{выход} \leq 1$	положить выход равным 1

Как это и бывает в действительности, в выборке представлены, главным образом, фирмы, которые, по мнению экспертов, должны выжить, и поэтому есть опасность, что сеть получит недостаточно информации, относящейся к потенциальным банкротам. Чтобы уравнивать представительство обеих групп, во время обучения сети

компаний, испытывающие трудности (группа 2), подавались на вход чаще других. Никакой новой информации к базе данных не добавлялось, но уже такое корректирование приводит к правильному расположению гиперплоскостей в пространстве входов.

Из-за малого размера базы данных и неравномерного распределения образцов по группам не представляется возможным разбить случайным образом все данные на обучающее, подтверждающее и тестовое множества. Поэтому для получения достоверной оценки доли случаев неправильной классификации применялся метод «складного ножа»: из выборки по очереди выбрасывалось каждое наблюдение. Для оценки обобщающей способности классификатора применялся 59-кратный метод перекрестного подтверждения. Окончательная оценка (так называемая «leave-one-out cross-validation error») вычислялась путем суммирования ошибок, полученных при выбрасывании отдельных наблюдений. В данном случае мы получаем 59 сетей, обученных на 58 наблюдениях каждая. Прием повторного использования наблюдений, несмотря на вычислительные затраты, хорош тем, что позволяет максимально использовать имеющуюся информацию, — важное качество, когда число наблюдений невелико.

Ошибка 59-кратного перекрестного подтверждения

	Класс 0	Класс 1	Класс 2	Всего	Число весов
Число скрытых элементов					
1	35%	100%	100%	55.90%	8
2	10%	100%	100%	39%	15
3	32.50%	68.80%	100%	45.70%	22
4	32.50%	75%	100%	47.50%	29
5	27.50%	100%	66.70%	49%	36
6	37.50%	87.50%	100%	54.20%	43

Таблица 8.3. Ошибки перекрестного подтверждения

Данные о количестве правильных и неправильных классификаций для 5-2-1 сети, которая имела наименьшую общую ошибку (39%), приведены в табл. 8.4.

Все члены 2-й группы (высокий кредитный риск) и 1-й группы (пограничные случаи) были классифицированы неправильно. Конкретно, двое «банкротов» получили кредиты, а трем «жизнеспособным» было отказано. Общий результат довольно плох, однако нужно отметить, что правильная классификация члена 2-й группы, пропущенного при обучении, представляет собой непростую задачу, по-

сколько остается всего два наблюдения для идентификации класса «банкротов».

Класс				Число неправильных классификаций
0	1	2		
0	36	1	3	10%
1	14	0	2	100%
2	2	1	0	100%

Таблица 8.4. Таблица ошибок перекрестного подтверждения для 5-2-1 сети

Все 59 обучающих множеств подавались на вход 4000 раз (с начальным состоянием датчика случайных чисел, равным 2). Результаты приведены вместе с соответствующими доверительными интервалами, потому что на компьютере Convex перекрестное подтверждение производилось с автоматически фиксированными начальными весами. При различных инициализациях могут получаться разные результаты, и таким образом можно обходить локальные минимумы. Уже первые эксперименты показали, что окончательный ответ очень чувствителен к изменениям установки начальных значений. Чтобы получить более надежные результаты, при обучении нужно выполнить несколько прогонов на 59-элементном множестве и затем оставить только те из них, которые соответствуют наилучшим индивидуальным результатам обучения (наименьшей средне-квадратичной ошибке на обучающем множестве).

Кроме того, мы брали сети с разным числом скрытых элементов (от одного до шести) и обучали их на всем обучающем множестве (метод повторной подстановки). Как и следовало ожидать, с увеличением сложности сети (числа весов) ошибка классификации уменьшалась, а для сети с шестью скрытыми элементами, вообще, оказалась равной нулю (полное обучение). Однако результаты перекрестного подтверждения не внушают доверия к надежности такого метода. Анализ результатов для модели 5-2-1 будет проведен ниже.

А сейчас мы хотим привести матрицы ошибок классификации для 5-2-1 модели и для MDA-метода (см. табл. 8.6).

Выбранная конфигурация сети дает приблизительно 80-процентную точность классификации — вдвое лучше, чем MDA. Это говорит об отличной способности сети к обобщению (вспомните, что наилучшие результаты сеть показала при перекрестном подтверждении). Особенно важно, что все три «промаха» были поняты. Таким образом, ошибки 1-го рода сведены к нулю. Следует помнить, что минимизация этих ошибок имеет смысл только при том условии, что не становятся слишком большими ошибки 2-го рода. Действи-

тельно, когда выборка состоит преимущественно из «жизнеспособных» компаний, даже малое количество ошибок 2-го рода сильно скажется на общей ошибке. По этому показателю сеть также опережает MDA.

Неправильные классификации на всем обучающем множестве					
	Класс 0	Класс 1	Класс 2	Всего	Число весов
Число скрытых элементов					
1	25%	81.30%	0%	39%	8
2	2.50%	68.70%	0%	20.30%	15
3	0%	43.70%	0%	11.90%	22
4	0%	12.50%	0%	3.40%	29
5	0%	6.25%	0%	1.70%	36
6	0%	0%	0%	0%	43

Таблица 8.5. Результаты обучения для сетей возрастающей сложности

Предсказанный номер группы MDA				
Действительный номер группы	Число случаев	0	1	2
Группа 0	40	16 40%	12 30%	12 30%
Группа 1	16	4 25%	6 38%	6 38%
Группа 2	3	0 0%	1 33%	2 67%
Доля правильно классифицированных случаев:		41%		
Предсказанный номер группы 5-2-1				
Действительный номер группы	Число случаев	0	1	2
Группа 0	40	39 97%	1 3%	0 0%
Группа 1	16	11 69%	5 31%	0 0%
Группа 2	3	0 0%	0 0%	3 100%
Доля правильно классифицированных случаев:		80%		

Таблица 8.6. Метод повторной подстановки. Результаты классификации для 5-2-1 модели и MDA

Убедившись в высоком качестве работы сети, мы обратились к исследованию поведения отдельных переменных. К сожалению, в данном случае каждая степень свободы отвечает за несколько переменных (это относится и к MDA), и роль каждой из них в отдельности с трудом поддается изучению. Поэтому мы заново обучили сеть на исходных, большей частью качественных, данных.

Результаты для 26-мерной входной матрицы

При столь небольшом числе степеней свободы (размерность равна 26) едва ли возможно построить сложную сеть, имеющую несколько (более одного) скрытых элементов и обладающую хорошими способностями к обобщению. Поэтому мы остановились на конфигурации 26-0-1, которая аналогична случаю логистической регрессии, и обучали сеть в течение 200 эпох. Получившийся в результате вектор весов приведен в приложении 3.

Один весовой коэффициент сам по себе мало что говорит о конечном вкладе переменной в решение. Разные комбинации входных значений по-разному взаимодействуют с комбинациями весов, поэтому при анализе необходимо учитывать и сами входные значения. Мы измеряли чувствительность начального решения к изменениям одной входной переменной. Этот метод дает возможность количественно оценить вклад каждой переменной в ответ для данного наблюдения. Вся последовательность действий такова:

- 1 Все значения одной из входных переменных x заменяются на ее безусловное ожидание (т.е. среднее арифметическое).
- 2 Выход сети перевычисляется для измененной таким образом входной матрицы при сохранении прежней матрицы весов.
- 3 Сравниваются друг с другом абсолютные остатки для такой входной матрицы и для исходной. Именно:

$$\text{Вклад переменной } x = \left| \frac{(\text{цель} - \text{прогноз}_{\text{новый}})}{(\text{цель} - \text{прогноз}_{\text{старый}})} \right|,$$

где

вклад < 0.5	влияние переменной сильно отрицательное
0.5 < вклад < 0.75	влияние умеренно отрицательное
0.75 ≤ вклад ≤ 1.25	влияние нейтральное
1.25 ≤ вклад ≤ 2.00	влияние умеренно положительное
вклад > 2.00	влияние сильно положительное

- 4 Прodelать действия из предыдущего шага для остальных 25 переменных.

В приложении 4 приведены данные частотного анализа индивидуальных вкладов по каждой переменной. Величина вклада < 0.75 оз-

начает, что эта переменная «завышает» оценку фирмы, которая по классификации была, скорее всего, отнесена к 1-й или 2-й группам (с высоким кредитным риском). Наиболее значительный вклад в результат классификации дают следующие 7 переменных:

A3	Положение на рынке
A4	Экспорт
B9	Благонадежность управляющего
C5	Зависимость от косвенных налогов
D1	Чистая маржа
D2	Коэффициент покрытия долга
D3	Отношение ликвидности

Все семь переменных влияют на то, в какую группу попадет данный элемент, при этом высокий кредитный риск связан, прежде всего, с экспортом и покрытием долга, тогда как низкий кредитный риск коррелирует со всеми семью переменными.

Коль скоро переменная, соответствующая кредитному риску, позволяет различать между собой группы объектов, то естественным следующим шагом представляется предложенный Горманом и Сейновским кластерный анализ весов-состояния (multivariate analysis along the lines, см. гл. 4). Далее, на решение оказывают влияние как финансовые, так и нефинансовые факторы. Было бы интересно разделить влияние этих двух типов переменных и оценить маргинальный вклад качественных переменных.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Эксперимент, где в качестве критерия было взято мнение группы экспертов в области кредитного дела, показал, что нейронные сети являются многообещающим средством для классификации *наперед* (прогнозирования). Это особенно важно для государств, где права кредиторов при банкротстве постоянно изменяются и где выживание фирмы сильно зависит от возможности получения кредитов. На практике задача построения нужной базы данных, состоящей, в основном, из качественных переменных, может оказаться непростой, однако при недостатке надежных количественных данных кредитные менеджеры так или иначе вынуждены оценивать кредитную привлекательность фирм, основываясь при этом на той, пусть неполной, информации, которой они располагают.

Разрабатывая нейронно-сетевой классификатор, мы пытались построить метод, позволяющий извлекать знания экспертов из сделанных ими оценок. Выяснилось, что ключевое значение имеют 7 переменных, среди которых есть и качественные, и количественные.

Было замечено, что результаты очень чувствительны к выбору начального приближения и потому могут не быть оптимальными. Для удобства мы брали все время одинаковое число эпох и одно и то же начальное приближение для весов. Значимые переменные были выделены по результатам исследования зависимостей по отдельным переменным, хотя более правильным в данном случае представлялся бы многомерный подход.

Мы не пытались применять в этой задаче нелинейный (например, квадратичный) MDA-анализ — вероятно, это была бы более подходящая точка отсчета для оценки качества нейронного классификатора. Разница между линейной и квадратичной моделью может сказываться в случае, когда точки попадают на участок насыщения логистической кривой. Также не рассматривались и методы k ближайших соседей ($kNN = k$ Nearest Neighbour). Они являются непараметрическими методами классификации и не требуют предварительного задания вида зависимости. Следовательно, такие методы могут давать лучшее качество аппроксимации на выборке за счет разбиения пространства переменных на произвольное число областей для последующего принятия решений. Однако, поскольку kNN -методы обычно разбивают базу данных с помощью расстояния Махаланобиса (d^2), при этом неявно предполагается, что межгрупповые матрицы дисперсии одинаковы (см. [19]). Более того, осмысленный прогноз возможен только в предположении, что эти матрицы постоянны во времени. В этом вопросе необходимы дальнейшие исследования.

ОПЫТ ОЦЕНКИ КРЕДИТНОГО РИСКА В ГОЛЛАНДСКОМ ИНВЕСТИЦИОННОМ БАНКЕ

Выше мы рассказали о результатах классификации гипотетической совокупности компаний на основании оценок, сделанных группой польских экспертов в области кредитов. Нельзя при этом исключить, что работники банков сочтут построение достаточно содержательной базы качественных данных слишком трудоемким делом, и весь метод, будучи теоретически правильным, может оказаться непрактичным. Частично такие сомнения снимаются тем обстоятельством, что уже в настоящее время Голландский инвестиционный банк (NIB) (торговый банк, специализирующийся на средне- и долгосрочном финансировании корпораций) использует порядка 44 качественных показателей для оценки кредитного риска по реальному инвестиционному портфелю, состоящему из займов, выданных банком своим корпоративным клиентам. Наряду с качественными показателями используются также 5 количественных. Используя информацию об этой работе, можно получить обобщения построенных нейронно-сетевых методов. При этом отсутствует возможность для

ex post-оценки ошибок 2-го рода, поскольку данные о клиентах, чьи заявки на получение займа были отвергнуты, не хранятся.

ОПИСАНИЕ БАЗЫ ДАННЫХ ГОЛЛАНДСКОГО ИНВЕСТИЦИОННОГО БАНКА

По соображениям коммерческой тайны банк уклонился от того, чтобы полностью раскрыть схему выбора и построения совокупности 44 качественных переменных. В общих словах, в качестве переменных были выбраны показатели успеха, которые были признаны решающими на основании 45-летнего опыта работы в сфере инвестиций. Данное множество переменных годится для описания малых и средних компаний, действующих в самых разных отраслях, но не подходит для многонациональных компаний и для долгосрочного финансирования промышленных проектов. Конкретнее, охватывающиеся следующие 5 групп факторов, влияющих на шансы фирмы выжить:

- 1 *Состояние рынка* (10 переменных). Сюда входит экзогенное окружение, на которое сама фирма влиять не может, в частности, влияние фазы торгового цикла и тенденции в структуре потребления.
- 2 *Положение на рынке* (9 переменных). Конкретные условия рынка, в которых действует данная фирма, анализируются более подробно в соответствующих терминах.
- 3 *Стратегия* (9 переменных). Цели компании на ближайшие годы с точки зрения их выполнимости и последовательности.
- 4 *Эффективность* (9 переменных). Производство или деятельность в сфере услуг оцениваются с точки зрения эффективности, логистики и дружелюбности окружения.
- 5 *Управление* (7 переменных). Работа высшего управляющего звена фирмы в прошлом и в настоящее время с учетом фаз жизненного цикла компании.

Все переменные принимают значения от 0 до 4, при этом соответствие следующее:

- | | |
|---|---------------------------|
| 0 | значительно ниже среднего |
| 1 | ниже среднего |
| 2 | на среднем уровне |
| 3 | выше среднего |
| 4 | значительно выше среднего |

Совокупность переменных дополняется пятью количественными показателями. Сюда входят: степень прибыльности (отношение операционного дохода к используемому капиталу), коэффициент платежеспособности (отношение собственного капитала к суммарному), ликвидность (отношение текущих активов к обязательствам), коэффициент покрытия (отношение операционного дохода к заемному капиталу) и разница между наивысшим и наименьшим показателями прибыльности за последние 4 года.

Итоговый показатель рисковости кредита получается как взвешенное среднее всех переменных. Он может принимать такие значения:

- 1 банкроты
- 2 перспективы неясны
- 3 жизнеспособные

По каждому из трех классов мы получили доступ к данным, позволяющим вычислить рейтинги 10 компаний. Равномерность распределения наблюдений по классам повышает представительность данных и облегчает обучение (сравните это с предыдущими задачами, где нам приходилось уравнивать неравномерность представительства в двух группах из трех). Так как особенности отраслей уже учтены в переменных, данные по разным группам не сравнивались между собой на предмет стандартизации факторов, характеризующих специфику отраслей и окружения.

Год оценки	Класс целевой переменной			всего
	плохо (1)	средне (2)	хорошо (3)	
1986	1	0	2	3
1987	4	0	1	5
1988	4	4	7	15
1989	1	4	0	5
1990	0	2	0	2
Всего	10	10	10	30

Таблица 8.7. Кросс-таблица классов целевой переменной по годам оценки

Рейтингам, полученным из качественных переменных, придается в банке большое значение. Оценка компании продлевается финансовым менеджером совместно со специалистами из отделов финансирования корпораций и финансового анализа. Несмотря на то, что переменные по своей сути в высшей степени субъективны, выработка единого мнения как минимум трех специалистов такую субъективность уменьшает. Разумеется, и это мнение может быть результа-

том коллективной ошибки (и такой пример будет далее приведен). Вся описанная довольно сложная процедура применялась совершенно одинаковым образом ко всем 30 компаниям из выборки при принятии решения о предоставлении им кредита в период с 1986 г. по 1990 г. Таким образом, значения переменных отражали мнение банка о перспективах заемщика на момент рассмотрения его просьбы о займе. После этого состояние дел заемщика ежегодно оценивалось вновь. Все 30 компаний находились в 3-й группе в течение, по меньшей мере, одного года после получения займа в период 1986–90 гг., так что для наблюдения *ex post* были доступны только ошибки 1-го рода. Поэтому в качестве *ex post*-оценивающей (целевой) переменной мы взяли результат переоценки от декабря 1992 г. В табл. 8.7 приведены частоты для целевой переменной по всем годам.

Анализ главных компонент

При наличии 49 переменных и 30 наблюдений возникает проблема, связанная с числом степеней свободы. Если число подлежащих оцениванию параметров превышает число наблюдений, то всегда можно достичь абсолютного соответствия модели наблюдаемым данным, но это соответствие будет на самом деле мнимым. Не имея возможности существенно увеличить число наблюдений, мы приняли решение уменьшить число переменных. При помощи нелинейного анализа главных компонент были выделены три наиболее значимых показателя (измерения), на которые приходится большая часть (около 60%) вариаций в исходной базе данных.

Измерение	Собственное значение
1	.2487
2	.1593
3	.1218

С первым измерением тесно связаны восемь переменных, причем все они являются эндогенными, т.е. зависят от способа управления, и характеризуют рыночную интуицию, приоритеты и дальновидность управления. Можно сказать, что это измерение служит количественным выражением качества управления фирмой. Следующее измерение связано с состоянием рынка вообще и положением данной фирмы на рынке и характеризует, в какой степени компания в состоянии должным образом реагировать на изменения обстановки. Третий показатель связан с эффективностью торговой политики фирмы. Производство с небольшой капиталоемкостью обычно в большей степени ориентировано на рынок в целом, а с большой капиталоемкостью — на продвижение конкретного продукта, и наоборот. Следует отметить, что количественные переменные модели вносят в эти

три измерения лишь относительно небольшой вклад. Малость этого влияния подтверждает предположение Арженти [15] о том, что финансовые переменные выражают скорее симптомы, чем причины явлений. В то же время исследуемые данные относятся к компаниям, чьи просьбы о предоставлении кредита были удовлетворены, так что с финансовыми показателями у них должно было быть все в порядке. Поэтому вопрос о том, существенно ли количественные показатели влияют на результат, остается открытым.

ДВЕ ТОЧКИ ОТСЧЕТА: MDA И kNN

Имея трехмерные числовые показатели, мы воспользовались линейным множественным дискриминантным анализом. Мы выбрали одинаковые (33%) априорные вероятности того, что данное наблюдение принадлежит той или иной группе. Данные о неправильных классификациях приведены в табл. 8.8.

Действительный номер группы	Число случаев	Предсказанный номер группы		
		1	2	3
Группа 1	10	6 60.0%	3 30.0%	1 10.0%
Группа 2	10	3 30.0%	4 40.0%	3 30.0%
Группа 3	10	0 0.0%	2 20.0%	8 80.0%
Доля правильно классифицированных случаев:		60%		

Таблица 8.8. Результаты классификации для MDA

Оценивая результаты классификации по методу MDA, мы видим значительную долю ошибочных решений по компаниям-банкротам (группа 1) — одной из них кредит был бы предоставлен. Фирмы с неясным положением (группа 2) с трудом поддаются правильной классификации, потому что, в конечном итоге, они могут попасть в 1-ю или 3-ю группу. Дело нельзя улучшить, приводя априорные вероятности в соответствие с представлениями банка о вероятности принадлежности фирмы различным группам. Общий показатель правильности прогноза составил всего 56.6%, причем из 1-й группы правильно классифицированы были только 30%.

В отличие от метода MDA, который разделяет все множество на группы линейно, метод k ближайших соседей (1-NN) выделяет группы, линейно разделяя их попарно. Коль скоро мы допускаем, что на-

ша задача в какой-то степени нелинейна, этот метод представляется подходящим. Результат аппроксимации выборочного распределения (с применением обычного евклидова расстояния) этим непараметрическим методом может быть лучше, чем в предыдущем случае. Однако приведенные в табл. 8.9 результаты классификации говорят об обратном. Возможно, более подходящими для данной задачи оказались бы NN-методы более высокого порядка (например, 3-NN).

		Предсказанный номер группы 1-NN		
		1	2	3
Действительный номер группы	1	3 30%	5 50%	2 20%
	2	3 30%	3 30%	4 40%
	3	0 0%	2 20%	8 80%
Доля правильно классифицированных случаев:		47%		

Таблица 8.9. Результаты классификации для метода ближайших соседей

РЕЗУЛЬТАТЫ КЛАССИФИКАЦИИ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Были опробованы различные нейронные классификаторы, основанные на стандартной MBPN-модели, и их результаты сравнивались с тем, что дают методы MDA и 1-NN. Чтобы сохранить естественное отношение упорядоченности между тремя классами, выход был выбран одномерным. Описанные выше три агрегированных количественных показателя никаким преобразованиям не подвергались, поскольку их значения лежат в интервале от -2 до 2 , и это вполне подходит для подачи на входной слой сети. Оба скрытых элемента и выходной элемент имели сигмоидную функцию преобразования с единичной крутизной и коэффициентом обучения 0.1 . Полученные на выходе значения разбивались на три группы в соответствии с предполагаемым вариантом ответа на основной вопрос:

- $0 \leq \text{выход} \leq 0.33$ положить выход равным 0 банкроты
- $0.33 < \text{выход} < 0.66$ положить выход равным 0.5 неясно
- $0.66 \leq \text{выход} \leq 1$ положить выход равным 1 жизнеспособные

Здесь, как и в предыдущих случаях, мы применяли процедуру перекрестного подтверждения (по очереди отбрасывая каждое наблю-

дение) с целью выбрать наилучшую конфигурацию сети (с точки зрения способности к обобщению) и получить оценку надежности модели. Каждый вариант конфигурации сети обучался в течение 30 циклов (3000 эпох) на 29 образцах, а 30-й образец оставлялся для проверки. Выбор подходящей архитектуры сети осуществлялся на основании статистики по среднеквадратичной ошибке (RMSE) перекрестного подтверждения (CV = Cross-Validation).

Тип сети	CV RMSE	Доля правильных прогнозов (%)	Число весов
3-0-1	0.42	50.00	4
3-1-1	0.40	56.70	6
3-2-1	0.45	40.00	11
3-3-1	0.58	36.60	16

Таблица 8.10. Данные перекрестного подтверждения для различных вариантов архитектуры сети

Была сделана попытка выбрать степень сложности модели, исходя из другого критерия — итоговой ошибки прогноза (FPE = Final Prediction Error). Эта величина вычисляется по ошибке на всем обучающем множестве с добавлением штрафного слагаемого за сложность модели:

$$FPE(\lambda) = MSE(\lambda)[(1 + S(\lambda)/N)(1 - S(\lambda)/N)], \quad (2)$$

где N — объем выборки, $S(\lambda)$ — число весов модели (λ).

Для линейных моделей в предположении, что объем выборки достаточно велик, этот критерий дает несмещенную оценку риска обобщения при прогнозе. Это утверждение верно в асимптотическом смысле при $N \rightarrow \infty$, и наши результаты указывают на то, что при $S(\lambda) \rightarrow N$ оно не выполняется. Утанс и Муди [270] утверждают, что несмещенные оценки могут быть получены также для нелинейных моделей (в частности, нейронных сетей).

Мы подсчитывали FPE для различных сетей возрастающей сложности, и в табл. 8.11 приведены результаты, соответствующие двум видам выходных элементов: многопорогового (0, 0.5, 1) и сигмоидного.

Для исследуемой задачи модель FPE1 (сигмоидная) представляется не вполне подходящей, так как FPE здесь более естественно было бы вычислять «ступенчато» (как это делается в FPE2). Критерий FPE2 имеет локальный минимум для конфигурации 3-1-1, что согласуется с результатами перекрестного подтверждения. Это, однако, говорит о том, что ни один из методов выбора модели не является идеальным (при конечном числе точек наблюдения). Поэтому мы остано-

вим свой выбор на перекрестном подтверждении как на наиболее надежном методе выбора варианта модели и рассмотрим 3-1-1 сеть.

Из таблицы результатов классификации для 3-1-1 сети (см. табл. 8.13) видно, что 80% случаев из 1-й группы и 90% из 3-й группы были классифицированы правильно. С другой стороны, из сомнительных компаний ни одна не была идентифицирована правильно.

Тип сети (*)	RMSE1	FPE1	RMSE2	FPE2
3-0-1	0.302	0.119	3.87	19.6
3-1-1	0.286	0.122	3.46	18.0
3-2-1	0.134	0.134	3.46	25.9
3-3-1	0.076	0.076	2.23	16.4
3-4-1	0.066	0.066	0	0

(*) Эти данные соответствуют фиксированному количеству эпох обучения (20 тыс.). Поскольку при дополнительном обучении среднеквадратичная ошибка MSE 3-4-1 сети уменьшится, результаты не носят абсолютный характер. RMSE1 соответствует сигмоидному выходному узлу, а RMSE2 — пороговому.

Таблица 8.11. Данные об итоговой ошибке прогноза

Сеть	Доля правильных прогнозов (%)	Число весов
1-NN	46.60	**
MDA	60.00	12
Сеть 3-0-1	60.00	4
Сеть 3-1-1	70.00	6
Сеть 3-2-1	70.00	11
Сеть 3-3-1	90.00	16
Сеть 3-4-1	100.00	21

Таблица 8.12. Метод повторной подстановки. Результаты для сетей возрастающей сложности

Мы проводили также обучение этих сетей на всем обучающем множестве (метод повторной подстановки), меняя при этом число скрытых элементов от 0 до 4. Как и следовало ожидать, при увеличении сложности сети (числа весов) ошибка классификации уменьшалась, а для сети с четырьмя скрытыми элементами даже было достигнуто полное соответствие в отображении (ошибка равна нулю). Так как для отделения друг от друга трех групп понадобилось 4 скрытых элемента, задача, очевидно, является нелинейно отделимой. Однако, результаты перекрестного подтверждения заставляют сомневаться в возможностях такой сети к обобщению. Далее будут анализировать-

ся результаты, полученные для модели 3-1-1 после 7000 эпох обучения.

Конфигурация 3-1-1 дает точность классификации примерно 70%, что на 10% лучше, чем MDA. В первой группе число правильно классифицированных элементов стало больше на 30% — прекрасный результат с учетом того, что структура базы данных позволяет анализировать только ошибки 1-го рода. Показательно, что и MDA, и нейронная сеть одновременно неправильно классифицировали одну из компаний 1-й группы (№13 в списке) как жизнеспособную (отнесли ее к 3-й группе). Такое совпадение заставило банк пересмотреть ее рейтинги. Оказалось, что эта компания получила очень высокие оценки по многим параметрам благодаря тому, что в тот момент качество оценки было несовершенным.

Предсказанный номер группы
3-1-1

Действительный номер группы	Число случаев			
		1	2	3
Группа 1	10	9	0	1
Банкроты		90%	0%	10%
Группа 2	10	5	2	3
Неясные случаи		50%	20%	30%
Группа 3	10	0	0	10
Жизнеспособные		0%	0%	100%
Доля правильно классифицированных случаев:		70%		

Таблица 8.13. Метод повторной подстановки. Результаты классификации для 3-1-1 сети

Нейронная сеть не смогла превзойти результаты метода MDA на 2-й группе компаний. Эта группа состоит из фирм с неясным будущим. В данный момент их финансовое положение неустойчиво, и, возможно, они справятся с этими трудностями, а, может быть, и нет. Было бы интересно сравнить результаты классификации с тем, что в действительности произошло с этими фирмами спустя некоторое время. Опять-таки, приятно отметить, что компании, неправильно классифицированные методом MDA, были также неправильно оценены нейронным классификатором. 100-процентный результат, достигнутый на 3-й группе, говорит о том, что эта часть компаний образует отчетливо выделяемую область в пространстве переменных.

Все полученные результаты собраны на рис. 8.1, где для каждой компании наряду с ее истинным номером группы (целевой переменной)

изображены результаты ее классификации 3-1-1 нейронной сетью, методами MDA и 1-NN. Для удобства дискретные выходные значения представлены в непрерывном виде.

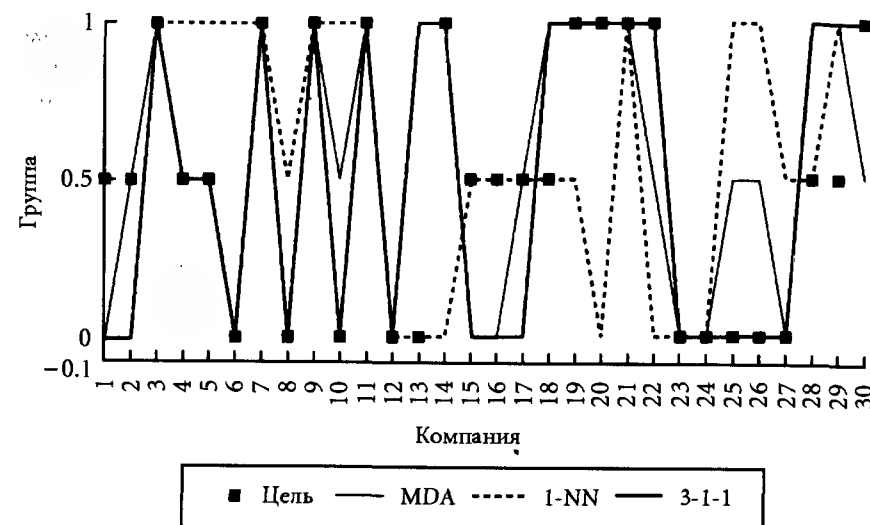


Рис. 8.1. Метод повторной подстановки. Результаты классификации различными методами

ОБСУЖДЕНИЕ

Голландским инвестиционным банком была разработана и использовалась реальная качественная база данных для оценки будущего состояния дел его корпоративных клиентов. Из-за того, что целевая переменная определялась задним числом, данные не позволяли исследовать ошибки 2-го рода. При помощи нелинейного анализа главных компонент первоначальное количество переменных (49) было уменьшено до 19 переменных, которые располагались в трех измерениях. В связи с тем, что, как выяснилось, около 60% исходных переменных несущественны для принятия решения, банк в настоящее время подумывает о том, чтобы изменить способ оценки. Как вариант рассматривается двухступенчатая процедура, в которой клиент сначала проходит сканирование по 19 переменным, а в случае успеха включаются в работу 30 новых переменных. Благодаря такому «двухъярусному» методу будут высвобождены значительные управленческие ресурсы банка. Малая степень значимости пяти финансовых показателей в полученном трехмерном пространстве может быть связана с тем, что в исследуемой выборке были представлены только жизнеспособные компании.

На вход нейронного классификатора, показавшего наилучшую способность к обобщению (таким оказалась 3-1-1 сеть), была подана часть образцов. Результаты классификации (70-процентное попадание) выглядят обнадеживающе на фоне традиционных методов — линейного множественного дискриминантного анализа (MDA) и метода ближайшего соседа (1-NN). Однако способность сети к прогнозированию (56.7% правильных ответов) оказалась ниже ожидаемой. Можно предположить, что целевая переменная была не вполне правильно специфицирована для сомнительных компаний (пограничных случаев), и истинную картину можно будет установить, сравнивая результаты, полученные сетью, с реальными показателями в будущем.

Тем не менее, результаты наших исследований побудили Отдел корпоративного финансирования банка к продолжению исследовательских работ в области извлечения экспертных знаний.

ПРИЛОЖЕНИЕ 1: ИССЛЕДОВАННЫЕ ФИНАНСОВЫЕ ОРГАНИЗАЦИИ

- Bank Handlowy, Варшава
- * Bank Inicjatyw Gospodarczych, Варшава
- * Bank Komercyjny 'Poznania', Познань
- Bank Plocki Spoldzielczy, Плоцк
- * Bank Pocztowy S.A., Быдгощ
- Bank Rozwoju Rolnictwa, Познань
- Bank Slaski S.A., Катовице
- Bank Spoldzielczy, Кутно
- BOS Bank, Варшава
- Citibank Poland S.A., Варшава
- Enterprise Credit Corporation, Варшава
- * Gliwicka Bank Handlowy, Гливице
- Gorno-laski Bank Gospodarczy S.A., Катовице
- Gospodarczy Bank Wielkopolski S.A., Познань
- Interbank S.A., Варшава
- Kredyt Bank S.A., Лодзь
- * Kredyt Bank S.A., Варшава
- National Bank of Poland, Варшава
- Pierwszy Polsko-Amerykanski Bank S.A., Краков
- Powszechny Bank Gospodarczy, Лодзь
- * Powszechny Bank Kredytowy S.A., Варшава
- * Societe Generale, Варшава
- * Wielkopolski Bank Kredytowy, Познань

* обозначены организации, в которых работал хотя бы один из 12 экспертов

1	2	3	4	5	6
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9
10	10	10	10	10	10
11	11	11	11	11	11
12	12	12	12	12	12

ПРИЛОЖЕНИЕ 2: АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТ: ВЕСА КОМПОНЕНТ

Измерение	Собственное значение	Смысл
1	.1766	макроэкономическая чувствительность
2	.1136	биография управляющего
3	.0989	рост объема продаж
4	.0958	отрасль промышленности
5	.0764	ликвидность

Переменная	Измерение				
	1	2	3	4	5
INDUSTRY	-.280	-.300	.301	-.665	.048
MARKPOT	.394	.176	.481	.170	-.021
POSMARK	-.308	.244	-.546	-.164	-.320
EXPORT	-.888	.052	-.126	.306	.055
PREMISES	.103	-.092	.327	.366	-.537
AGECOMP	.316	.122	-.204	-.198	-.016
OWNERSH	-.419	-.509	.118	-.507	.074
RELATION	-.127	-.158	-.244	-.225	-.513
CENTRDEC	.239	-.115	.378	-.192	-.484
FLEXIB	.101	-.047	.555	.181	-.009
NOEMPLOY	.044	.544	-.198	.594	-.051
AGEMAN	-.299	.801	.098	-.331	.059
EXPBUSIN	-.058	.856	.109	-.307	.063
EXPFIELD	-.256	.669	.176	-.482	.096
RELIAB	.301	.064	-.484	-.106	-.340
SENSEXCH	-.931	-.014	-.036	.170	-.115
SENSINFL	-.931	-.059	-.018	.147	.020
SENSCUST	-.798	-.009	.136	.301	-.093
SENSDIRT	.076	.450	.239	-.202	-.493
SENSINDT	.190	.211	.052	.539	-.067
ENVIRONM	-.300	-.197	.434	-.166	-.217
NMARGIN	-.450	.103	.009	.083	-.367
DEBTSERV	-.169	.058	.399	.106	.124
QUICK	.084	.118	.281	.104	.568
SALESGR	.186	-.073	-.604	-.127	.048
RECEIVAB	-.281	-.025	-.255	-.051	.400

ПРИЛОЖЕНИЕ 3: МАТРИЦА ВЕСОВ ДЛЯ 26-0-1 СЕТИ

Порог	3.596		
A4	0.910	B6	-0.085
A6	0.230	C1	-0.092
C6	0.1	A5	-0.107
C4	0.098	A1	-0.120
B1	0.056	B2	-0.125
A2	0.037	D4	-0.140
B8	0.006	B5	-0.170
D5	-0.010	A3	-0.190
B3	-0.030	C5	-0.230
B4	-0.041	D1	-0.240
B7	-0.041	B9	-0.300
C3	-0.042	D2	-0.310
C2	-0.080	D3	-0.390

ПРИЛОЖЕНИЕ 4: ЧАСТОТНЫЙ АНАЛИЗ ВКЛАДОВ ОТДЕЛЬНЫХ ПЕРЕМЕННЫХ

		<0.5	0.5 ≤ 0.75	0.75 ≤ 1.25	1.25 ≤ 2	>2
A1	INDUSTRY	4	6	34	11	4
A2	MARKPOT	2	1	49	5	2
A3	POSMARK	4	6	25	19	5
A4	EXPORT	11	12	7	14	15
A5	PREMISES	1	5	44	7	2
A6	AGECOMP	4	3	41	7	4
B1	OWNERSH	0	6	47	4	2
B2	RELATION	3	9	33	11	3
B3	CENTRDEC	1	2	54	2	0
B4	FLEXIB	2	0	51	4	2
B5	NOEMPLOY	5	3	34	12	5
B6	AGEMAN	2	2	49	4	2
B7	EXPBUSIN	0	2	53	2	2
B8	EXPFIELD	0	2	57	0	0
B9	RELIAB	8	7	17	16	11
C1	SENSEXCH	6	1	40	8	4
C2	SENSINFL	3	1	46	6	3
C3	SENSCUST	1	4	48	4	2
C4	SENSDIRT	4	5	40	6	4
C5	SENSINDT	7	6	23	14	9
C6	ENVIRONM	4	5	40	6	4
D1	NMARGIN	7	4	22	18	8
D2	DEBTSERV	10	8	12	18	11
D3	QUICK	6	9	16	14	14
D4	SALESGR	7	4	30	14	4
D5	RECEIVAB	0	0	58	1	0

9

Прогнозирование банкротства корпораций

В предыдущей главе были изложены результаты классификации «*ex ante*» (т.е. наперед) в задаче оценки привлекательности фирм с точки зрения размещения кредитов, которую выполняли несколько польских специалистов в области кредитного дела на материале гипотетических заявок на предоставление кредита; также были рассмотрены результаты классификации «*ex post*» (т.е. задним числом) на материале реального инвестиционного портфеля Голландского инвестиционного банка. В обеих задачах имеющихся данных было недостаточно для того, чтобы проанализировать цену ошибок классификации. В этой главе мы займемся оценкой ошибок классификации 1-го и 2-го родов в задаче прогнозирования банкротства корпораций и исследуем возможные результаты применения нейронных сетей. Материалом для конкретной задачи послужат данные по ряду британских компаний — производителей комплектующих для автомобилей. В основе конструкции сети лежат семь входных моделей, использующих показатели, которые входят в так называемое *дзета* Альтмана.

ВОЗМОЖНОСТИ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ БАНКРОТСТВА КОРПОРАЦИЙ

Во многих реальных задачах основной трудностью оказывается то, что нейронная сеть не может достаточно ясно показать причинно-следственные связи и выдает какое-то решение по принципу черного ящика. При этом в финансовом анализе для оценки состояния дел предприятий уже давно используются специально подобранные комбинации различных показателей, а качество модели оценивается с помощью критериев согласия без учета структуры модели. По существу, все сводится к выбору показателя (или комбинации показателей), соответствующего решающему правилу, которое позволяет включить (или не включать) данное предприятие в ту или иную группу (жизнеспособные, быстро растущие, высокоприбыльные).

В принципе, данные по компаниям могут служить основой для нейронных сетей — очевидно, что здесь задействованы процессы,

близкие к случайному блужданию. Это не покажется удивительным, если посмотреть, как компании почти ежедневно взаимодействуют с рынком, конкурентами и постоянно меняющимися условиями работы, а также если учесть, что компания, способная сделать удачный прогноз на основе опережающей информации, имеет на финансовом рынке высокую ликвидность своих акций и большие возможности для усиления своих позиций.

При имеющемся уровне сложности и одновременности происходящих процессов модели, основанные на причинных связях, имеют ограниченные возможности для применения: вновь происходящие события постоянно меняют спецификации всех переменных (и включенных, и не включенных в модель), а значения априорных вероятностей и размеров выплат по различным стратегиям весьма неопределенны и резко меняются вместе с изменениями показателей экономического роста, процентных ставок, обменных курсов и прибыльностью сделок, не связанных с кредитованием (например, при изменении операционных и комиссионных сборов).

Традиционный подход к прогнозированию банкротств основан на множественном дискриминантном анализе (см. [7], [10], [12], [32], [33]). Методы такого типа используются в широко распространенных системах определения рейтинга кредитоспособности, где ищется гиперплоскость, наилучшим образом разделяющая «хороших» и «плохих» кандидатов. Хотя к настоящему времени разработано множество дискриминантных моделей, используется (в частности, в управлении кредитами) лишь небольшое число из них. В ряде случаев банки приходят к выводу, что методы MDA не дают ожидаемого улучшения точности по сравнению с традиционными методами.

Оценить качество моделей этого типа непросто, поскольку даже если погрешность вычисляется на материале новых наблюдений, данные по обанкротившимся компаниям (так же, как и по их выжившим партнерам) приходится большей частью на периоды интенсивных банкротств, и количество наблюдений, соответствующих выжившим компаниям, как правило, очень мало и дает лишь поверхностное представление о пересечении множеств банкротов и небанкротов. Поэтому надежность моделей MDA в реальном времени остается под вопросом. В принципе, нейронные модели справляются с некоторыми из этих трудностей, потому что для обучения могут быть взяты данные, охватывающие периоды с различной ситуацией в экономике и степенью интенсивности банкротств, и обанкротившиеся и выжившие компании могут быть представлены в обучающем множестве в пропорции, соответствующей реальной вероятности того и другого. К тому же сеть может учитывать большое количество разнотипных переменных и обучаться на них на всех, тогда

как данные для множественного дискриминантного анализа редко включают в себя экономические и финансовые индикаторы.

ОЦЕНКА КАЧЕСТВА МОДЕЛИ

Оценка качества моделей классификации представляет собой сложную задачу, потому что в большинстве реальных приложений цена ошибок неодинакова. Так например, отказ в кредите хорошему клиенту влечет за собой лишь организационные расходы на поиск нового клиента, тогда как предоставление кредита ненадежному партнеру может привести к большим убыткам. Из-за этой несимметрии денежных потоков при определении степени точности модели необходимо учитывать последствия того или иного прогноза. Качество прогнозирования банкротств определяется и тем, насколько точно выявляются банкроты, и тем, насколько точно классифицируются небанкроты. Необнаружение компании-банкрота называется ошибкой 1-го рода, а прогноз банкротства, которого на самом деле не последовало, — ошибкой 2-го рода.

Из-за несимметрии цены ошибки становится чрезвычайно трудно найти компромисс между ошибками 1-го и 2-го родов. Говоря иначе, трудно ответить на вопрос, которая из двух моделей лучше: та, которая правильно идентифицирует 90% банкротств и на одну правильную классификацию дает 10 ошибок 2-го рода, или та, которая идентифицирует 80% банкротств, но дает только 8 ошибок 2-го рода на одну классификацию.

Как можно видеть из рис. 9.1, точки X и Y , показывающие качество прогнозов, соответственно, для моделей x и y , не дают возможности однозначно сказать, какая модель лучше. Если снизить требования к точности прогнозирования банкротства, то может оказаться, что модель x , по-прежнему, будет давать большее число ошибок 2-го рода и, тем самым, будет уступать модели y в таких приложениях, где ошибки 2-го рода относительно дороги по отношению к ошибкам 1-го рода.

Из сказанного следует, что качество модели прогнозирования банкротств можно оценить только при условии, что заранее заданы цена ошибок и вероятность банкротства/выживания. Если, например, нам известно, что цена одного не предсказанного вовремя банкротства равна цене пяти ложных тревог, и что потерпит банкротство один процент компаний, то мы можем оценить модели:

$$\text{Цена ошибки модели } x: 1\% \times \{5 \times (10\%) + 10 \times (1 - 10\%)\} = 0.095,$$

$$\text{Цена ошибки модели } y: 1\% \times \{5 \times (20\%) + 8 \times (1 - 20\%)\} = 0.074.$$

Модель x уступает модели y , которая пропускает 20% банкротов, но имеет более низкий показатель ошибок 2-го рода.

Если же одно пропущенное банкротство стоит 30 ложных тревог, то ошибки этих двух моделей будут такими:

Цена ошибки модели x : $1\% \times \{30 \times (10\%) + 10 \times (1 - 10\%)\} = 0.120$,

Цена ошибки модели y : $1\% \times \{30 \times (20\%) + 8 \times (1 - 10\%)\} = 0.132$,

и модель x оказывается лучше, чем y . Заметьте, что наибольший вклад в погрешность модели вносит большое количество ошибок 2-го рода, и так получается потому, что они совершаются на жизнеспособных компаниях, а таких — подавляющее большинство. Решить, достаточно ли существенно отличаются результаты обеих моделей, чтобы по ним можно было высказывать предпочтение, здесь довольно трудно, так как неизвестно, как модель x будет работать при каком-либо совсем другом соотношении между ценами ошибок. При сдаче экзаменов на аудитора экзаменуемые распознают только 25% компаний-банкротов, но зато на каждый правильный прогноз приходится только 4 ложных тревоги (см. [258]).

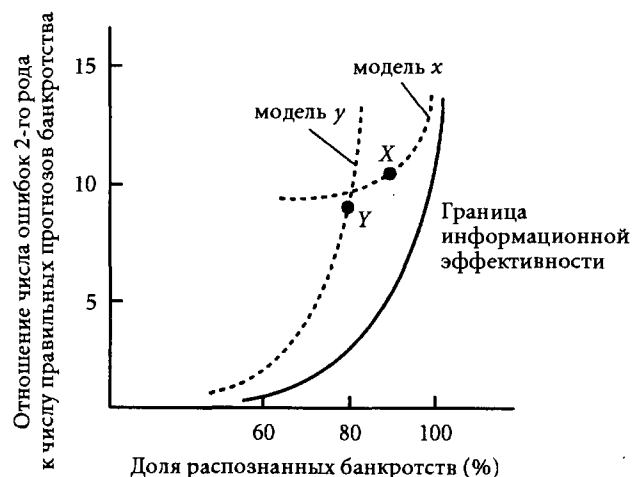


Рис. 9.1. Ошибки 2-го рода в прогнозе банкротств

При том, что MDA-модели способны распознавать гораздо большую долю компаний-банкротов, общий результат не обязательно будет лучше, чем практика аудиторов. Так, если цена одной ошибки 1-го рода в 5 раз больше, чем ошибки 2-го рода, то итоговая погрешность аудиторов будет равна

$$1\% \times \{5 \times (75\%) + 4 \times (1 - 75\%)\} = 0.0475,$$

что меньше, чем у обеих гипотетических моделей x и y .

Эти примеры иллюстрируют следующую мысль: если соотношения цен ошибок различаются сильно, то настройка модели банкротств на конкретные пропорции, по-видимому, более важна, чем

качество модели как таковое. Можно ввести обобщенное понятие информационной значимости модели, используя расстояние до так называемой эффективной информационной границы, т.е. кривой, огибающей результаты всех моделей. На рис. 9.1 модель x расположена ближе к этой границе, чем модель y , и поэтому ее следует считать более информационно эффективной.

Следующая проблема — это выработка стандарта для тестирования. Для оценки MDA-моделей в большинстве случаев берется небольшое количество образцов, и это увеличивает вероятность того, что модель будет слишком точно подогнана под тестовые данные. В выборках обычно содержится поровну компаний-банкротов и небанкротов, а сами данные, как правило, соответствуют периодам интенсивных банкротств. Это приводит к выводу о том, что надежными являются только результаты оценки модели на новых данных. Из табл. 9.1 видно, что даже на самых благоприятных тестах с новыми данными (когда все примеры берутся из одного периода времени и притом однородными в смысле отраслей и размера предприятия) качество получается хуже, чем на образцах, по которым определялись параметры модели. Поскольку на практике пользователи моделей классификации не смогут настраивать модель на другие априорные вероятности банкротства, размер фирмы или отрасль, реальное качество модели может оказаться еще хуже. Качество может также ухудшиться из-за того, что в выборках, используемых для тестирования MDA-моделей, бывает мало фирм, которые не обанкротились, но находятся в зоне риска. Если таких «с риском выживающих» фирм всего четыре-пять, то это искажает реальную долю рискованных компаний, и в результате частота ошибок 2-го рода оказывается недооцененной.

	Оценивание		Новые данные	
	Доля ошибок 1-го рода (%)	Доля ошибок 2-го рода (%)	Доля ошибок 1-го рода (%)	Доля ошибок 2-го рода (%)
Альтман [7]	6	3	4	21
Альтман [12]	3.8	10.3	7.5	10.3
Люерти (1989)	6.6	6.6	11.1	11.1

Таблица 9.1. Ошибки при определении параметров модели по выборке и на новых данных

ЭКСПЕРИМЕНТ

Пьесс и Вуд [217] провели сравнительную оценку Z-модели Альтмана с аналогами моделей «Datastream» и «Z Таффлера» (см. [256]), основываясь на точности классификации по данным временных ря-

дов, выражающих результаты деятельности компаний за период с 1973 по 1986 гг. Этот подход отличается от общепринятого, когда берутся данные за какой-то период времени, предшествовавший банкротству.

Высокая точность модели на выборочных данных, соответствующих компаниям, о которых уже известно, обанкротились они или нет, никак не поможет пользователю систем финансовой классификации делать деньги. Поэтому более правильным способом оценить предполагаемую точность модели представляется такой, где на протяжении определенного отрезка времени исследуются показатели ряда компаний, о которых неизвестно заранее, ждет ли их банкротство.

Поскольку новые данные для тестов берутся из той же совокупности, что и образцы, и имеют то же среднее значение, функцию распределения и частоту того или иного исхода, начинает вызывать сомнения надежность модели при использовании ее в реальном времени. Особенно плохо все становится в тех случаях, когда целевое состояние — банкротство корпорации, тяжелое состояние больного или обнаружение при проверке багажа спрятанного оружия — является весьма редким событием. Для банкротств корпораций это — порядка одного процента случаев. В такой ситуации даже очень точные модели при использовании в реальном времени выдают огромное количество ложных тревог. Так, например, доля ошибок в 10% при условии, что 99% компаний выживут, означает, что на каждую правильную идентификацию будет выдаваться примерно 10 ложных тревог (ошибок 2-го рода). Более того, редко происходящие события имеют большой разброс (дисперсию). Поэтому доля компаний, обанкротившихся в течение года, сильно меняется от года к году, а для небольших выборок, которые обычно являются основой базы данных банка или финансовой компании, этот эффект выражен еще сильнее.

Таким образом, для того чтобы получить хороший тест, нужно, прежде всего, учитывать все эти свойства, а уже потом — известные итоги работы компаний или среднюю предрасположенность фирм к банкротству. В эксперименте, о котором говорилось выше, тестовая база данных включала в себя информацию за 1973–86 гг. по всем компаниям-производителям комплектующих для автомобилей, акции которых котируются на финансовом рынке Великобритании. Из 24 компаний, действовавших в течение первого года этого временного отрезка, до самого его конца просуществовали 17. Четыре из оставшихся семи компаний подверглись слиянию, будучи вполне платежеспособными. Еще две компании попали под процедуру формальной ликвидации, а последняя была подвергнута реорганизации и реструктуризации под правительственным контролем, что эквива-

лентно банкротству. Таким образом, требовалось объяснить 258 «событий», из которых банкротств было три.

Три банкрота из 24 составляют 12.5%, что очень близко к предложенной Таффлером [256], а также Альтманом и др. [12] априорной доле банкротств в 10%. В то же время три факта банкротства на 258 событий дают интенсивность банкротств примерно на уровне одного процента, и это согласуется с годовой интенсивностью банкротств, зафиксированной по результатам наблюдений Дуна и Брэдстрита (1982).

РАЗРАБОТКА МОДЕЛИ

Модель Альтмана, которой мы пользовались, основывается на переменных и параметрах Z -модели 1968 года. У следующей версии модели — «Дзета» — были опубликованы только переменные, а параметры — нет. То же и для Z -шкалы Таффлера: переменные известны, а коэффициенты неизвестны, но могут быть приближенно восстановлены по опубликованным показателям шкалы. Наконец, шкала Datastream (прошлые показатели и данные по обанкротившимся компаниям были удалены из базы данных) приближалась с помощью переменной, которая, согласно публикации [257], сильно скоррелирована с Z модели Datastream.

В основу нейронной сети были положены семь входных переменных, базирующихся на показателях, входящих в Дзета-модель Альтмана. Поскольку банкротство — событие весьма редкое, а из компьютерной базы данных была исключена информация по обанкротившимся компаниям, мы не делали попыток обучить сеть на результат «банкротство/выживание». Вместо этого в качестве целевой переменной при обучении сети было взято состояние курса акций компании относительно общего индекса курсов акций. При этом мы исходили из того, что при ухудшении состояния дел компании ее акции падают в цене. Таким образом, выходной сигнал сети — не двоичный (0,1), а представляет собой переменную с непрерывно меняющимися значениями. Одновременно использовалась еще одна переменная, которая разделяла выход на банкротов и небанкротов таким образом, чтобы достигалась наилучшая относительная точность прогноза и относительная цена ошибок.

Для обучения сети были взяты данные по машиностроительным компаниям, акции которых котируются на фондовом рынке Великобритании. Обучающее множество состояло из данных по 20 компаниям за период времени с 1978 по 1986 гг. — всего 160 наблюдений. Модель представляла собой трехслойную сеть с семью входными элементами, соответствующими финансовым переменным. В скрытом слое было три элемента, и был также один выходной элемент.

Использовался метод обратного распространения ошибки с постепенным уменьшением коэффициента обучения по правилу Липмана, и модель достигала сходимости после 2000 итераций (эпох), при этом квадратный корень из среднеквадратичной ошибки составил 6.2% от среднего отклонения доходов.

СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Так как в реальной ситуации нельзя знать заранее, какая часть из компаний, представленных в случайной выборке, потерпит банкротство в течение года и поскольку авторы двух рассматриваемых моделей, как можно предположить, устанавливали разделяющие уровни, исходя из каких-то конкретных предположений об априорных вероятностях банкротства и цене ошибок, мы упростили процедуру сравнения и ввели относительные разделяющие уровни. Иначе говоря, для каждой модели мы считали сигналами о банкротстве нижние 10% сигналов, выдаваемых моделью за очередной год. На деле такой подход означает общую 10-процентную априорную вероятность банкротства и такое отношение числа сигналов о банкротстве к реальным банкротствам в предыдущем тесте, которое определяется с помощью оптимизирующего порога. Кроме того, этот способ имеет то преимущество, что при этом минимизируются искажения, возникающие из-за большого разрыва во времени между публикацией «Z-счета» Альтмана и проведением эксперимента. Средние показатели за это время могли измениться, и поэтому разделение компаний на сильные и слабые, исходя из определенной пропорции, представляется более надежным. В табл. 9.2 приведены результаты эксперимента по прогнозированию банкротств на год вперед с указанием погрешности для каждой модели.

	Z Альтмана	Аналог модели Z Таффлера	Аналог модели Z Datastream	Нейронная сеть
Правильная классификация				
Выживание	239	231	233	234
Банкротство	0	1	1	2
Неправильная классификация				
Выживание	19	27	25	24
Банкротство	3	2	2	1

Таблица 9.2. Точность различных прогнозов банкротства

Результаты говорят о том, что предсказание банкротства — сложная задача. Из-за того, что нормы отчетности весьма растяжимы,

вполне может получиться так, что в какой-то год фирма рапортует о прибылях, а в следующем году она банкротится. Компании также могут потерпеть банкротство из-за мошенничества или банкротства основного потребителя, и поэтому добиться высокого уровня точности непросто. При всем этом нейронная сеть показала лучшие результаты, чем другие модели: она обнаружила два из трех надвигающихся банкротств и неправильно классифицировала жизнеспособные компании в пропорции 12 на один правильный прогноз банкротства. В количественном выражении это в два раза лучше, чем аналоги моделей Таффлера и Datastream, а Z Альтмана, вообще, оказалась не в состоянии выявить ни одного банкротства.

Отсюда следует тот вывод, что во всякой классификации банкротства главной является надежность, и что было бы ошибкой оценивать качество модели по выборочным критериям, которые, скорее всего, нельзя использовать для расчетов наперед (*ex ante*). В частности, в задаче прогнозирования банкротств корпораций нейронная сеть показала на настоящих *ex ante*-тестах такой же уровень точности, какой дают MDA-модели на гораздо менее требовательных *ex post*-примерах методом исключения одного наблюдения из выборки.

ИСПОЛЬЗОВАНИЕ В НЕЙРОННОЙ СЕТИ ПОНИЖЕННЫХ РАЗДЕЛЯЮЩИХ УРОВНЕЙ

Участвовавшие в сравнении MDA-методы были рассчитаны и оптимизированы, исходя из доли ложных сигналов 10:1 при некоторых априорных вероятностях и цене ошибок. Хотелось бы использовать в качестве *ex ante* критерия меньшее, чем 10-процентное, число потенциальных банкротов в популяции, но это плохо согласуется с параметрами моделей. Это также противоречит практике, когда снижение порога ниже 10-процентного уровня не приводило к банкротству. Так, когда доля ложных сигналов урезалась до 7%, Z-шкала Таффлера вообще переставала идентифицировать банкротства, а модель Datastream наталкивалась на это препятствие на отметке 8%. В противоположность этому нейронная сеть распознала два случая банкротства ниже разделяющего уровня в 4.5%, т.е. сеть способна работать в условиях, когда на одну правильную идентификацию банкротства приходится всего пять ложных сигналов. Этот показатель сравним с наилучшими результатами, которые получаются у MDA-моделей на гораздо менее требовательных тестах задним числом (*ex post*). Отсюда следуют два вывода: во-первых, нейронные модели представляют собой надежный метод классификации в кредитной сфере, и, во-вторых, использование при обучении в качестве целевой переменной цены акции может оказаться более выгодным, чем собственно показателя банкротство/выживание. В цене акций отражает-

ся «эффективная» реакция на публикацию отчета о состоянии дел компании, тогда как банкротство/выживание, на самом деле, связано с решениями банка в вопросах кредитования, которые часто основаны на конфиденциальной информации, не отраженной в отчетах. Это приводит к довольно неопределенной и неустойчивой ситуации в принятии решения, поскольку получается, что банки фактически решают, что им делать в будущем, опираясь на свои прошлые решения!

10

Использование нейронных сетей в торговле

Главная задача инвестора — купить дешевле и продать дороже. Чем выше изменчивость цены актива, тем больше имеется возможностей для проведения выигрышных стратегий торговли, даже с учетом затрат на совершение сделок. К сожалению, то, что кажется простым и очевидным задним числом (*ex post*), бывает совсем неочевидно наперед (*ex ante*). Ключевой вопрос здесь заключается в том, можно ли определить направление, величину и волатильность будущих изменений цены актива, экстраполируя имеющиеся прошлые данные. Примеры, разобранные в этой главе, показывают, что адаптивные нелинейные системы могут быть обучены так, чтобы выполнять технический анализ при минимально возможных допущениях. Наш подход к задачам этого типа отличается от других тем, что мы не пытаемся оптимизировать входные данные. На вход сети подавались необработанные данные о ценах акций компании Юнилевер в прошедший период, а сеть была обучена на то, чтобы выдавать ежедневные распоряжения на совершение сделок.

ТЕХНИЧЕСКИЙ АНАЛИЗ И ГИПОТЕЗА ЭФФЕКТИВНОГО РЫНКА

В последнее время гипотеза эффективного рынка (ЕМН = Efficient Market Hypothesis), которую мы уже обсуждали в гл. 3, подвергается серьезной критике, и, как ни странно, эта критика исходит из академических кругов. В своей слабой форме эта гипотеза утверждает, что инвестор не может получить дополнительный доход (с учетом компенсации за риск, связанный с данной стратегией) за счет использования правил торговли, основанных на прошлых данных. Иными словами, информация о прошлых ценах и доходах не может принести пользу для извлечения дополнительного дохода. В то же время ЕМН-гипотеза не конкретизирует ни природу такой информации, ни способы ее извлечения из прошлых цен. Должна ли для этого использоваться обычная автокорреляция временных рядов, методы Бокса-Дженкинса или анализа Фурье, или какой-то из многочисленных методов фильтрации? Более того, ЕМН является *комбинированной гипотезой* в том смысле, что для ее проверки она требует предварительно-

го формирования модели ценообразования, которая, в свою очередь, зависит от степени прогнозируемости, и это еще более усложняет дело.

Причиной, не позволяющей отвергнуть гипотезу ЕМН, является присутствие на рынке фондов, работающих с индексами курсов акции, и иных форм пассивного менеджмента, которые особенно популярны среди пенсионных фондов (см. [99], [100]). Положение дел, когда судьба средств пенсионеров зависит, в основном, от усредненных показателей фондового рынка, — это торжество той точки зрения, согласно которой инвестиционное сообщество не обладает достаточной квалификацией для того, чтобы получать дополнительные доходы. Однако в последнее время в академических изданиях стало модным высказывать предположение, что финансовые рынки обладают некоторыми признаками прогнозируемости (см. [90], [216]). Мысль о том, что различные уровни волатильности цен собраны в кластеры во времени, привела к появлению моделей ARCH (= Auto Regressive Conditional Heteroschedasticity = авторегрессионная условная гетероскедастичность). Задолго до этого «ARCH-взрыва», еще в начале столетия, основоположник технической торговли Чарльз Доу сформулировал первоначальный вариант теории, носящей теперь его имя. Говоря вкратце, в техническом анализе (который также называют чартизмом; chart — карта, диаграмма) утверждается, что в диаграммах временных рядов цен содержится информация о том, как инвесторы реагируют на новые события. Понимание психологии рынка может помочь аналитику предсказывать будущие тенденции. Как показали Брок, Лаконишок и Ле Барон [56], при некоторых предположениях такие известные чартистские методы, как *правило превышения предела изменения цен* (TRB = trading-range break) и *правило скользящего среднего* (MA = moving-average), могут дать прибыльную стратегию торговли. Правило TRB говорит, что следует заключать сделки на покупку, когда цена превысит свое предыдущее наивысшее значение, и на продажу, когда цена упадет ниже последнего минимума, а правила MA основаны на том соображении, что следует делать покупки, когда краткосрочные скользящие средние превышают (пересекают) долгосрочные скользящие средние, и продавать, когда краткосрочные становятся ниже долгосрочных. Несколько вариантов этих правил были опробованы на данных об индексе Доу-Джонса (Dow Jones stock index) с 1897 г. по 1988 г., при этом в качестве исходных точек для сравнения использовались ряд, порожденный случайным блужданием, и модель GARCH. Оба правила торговли порождали существенные прибыли: за распоряжениями о продаже следовали падения цены в среднем на 9%, а за сигналами на покупку — повышения цены в среднем на 12% (из расчета за год). Ни одна из моделей, с которыми проводилось сравнение, ощутимых доходов не принесла.

К сожалению, успех в применении технического анализа полностью зависит от качества метода оптимизации, о которой говорилось выше. Взяв длину промежутка для скользящего среднего равной, например, 125 дням, мы тем самым неизбежно ограничиваем свой выбор среди различных характеристик временных рядов для данной базы данных. Следует понимать, что такие действия, не сопровождающиеся достаточно хорошим подтверждением, могут привести к переобучению и потере способности к обобщению. Более того, многие инвесторы считают, что ключом к успеху в инвестиционном деле является интуиция аналитика, а не применение какой-либо процедуры отбора или формулы. В связи с этим Холи и др. [136] утверждают, что хотя успехи нейронных сетей в распознавании образов и делают возможным их использование в техническом анализе, все же наиболее выигрышные приемы будут, скорее всего, разработаны самими чартистами. Высказываются также предостережения против чрезмерной предварительной обработки входных данных, и поэтому мы воздержались от сверхоптимизации данных. Для того чтобы сохранить статистическую представительность данных, мы выбрали для обучения сети недельный промежуток времени. Для 5-20-1 сети это дает примерно 10 наблюдений на один весовой коэффициент.

Вычисление технических индикаторов типа скользящих средних или индексов относительной силы следует рассматривать как методы распознавания образов. В принципе, нейронные сети способны классифицировать образы и, следовательно, предсказывать структуру доходов. В одной из первых работ в этой области, посвященной прогнозированию доходов по акциям IBM, Уайт [279] показал, что нейронные сети предоставляют еще один интересный способ проверки ЕМН-гипотезы. Если бы рынок был полностью эффективным, то доходы описывались бы случайным блужданием и, следовательно, были бы совершенно непредсказуемы. Гипотезу можно проверить, погружая данные о доходах в пространство задержек, которое определяется как совокупность наборов значений дохода, соответствующих настоящему времени и предыдущим моментам времени до заданного порядка p : $(r(t), r(t-1), \dots, r(t-p))$.

Непредсказуемость означала бы, что координаты точек временного ряда, соответствующие компонентам задержки, образуют случайное распределение вокруг гиперплоскости $r(t) = 0$, т.е. белый шум. Если же распределение не является случайным, то при помощи нейронной сети можно аппроксимировать среднюю поверхность, отклоняющуюся от гиперплоскости $r(t) = 0$.

В следующем разделе мы подробно рассмотрим данные по компании Юнилевер и правила торговли, которые были получены.

СБОР ДАННЫХ И ОПРЕДЕЛЕНИЕ ПРАВИЛ

Из базы данных Datastream мы выбрали ежедневные данные о ценах акций компании Юнилевер (в голландских гульденах) за период с января 1973 г. по март 1992 г., что в совокупности составило около 5000 записей. Затем мы разделили все данные на две части: первые 2500 записей использовались как обучающие множества для нейронно-сетевых моделей, а оставшаяся 2501 запись — для тестирования (см. табл. 10.1). Мы выбрали в качестве примера большую транснациональную компанию Юнилевер потому, что ее консервативность и устойчивость, отсутствие реорганизаций и поглощений приводят к стабильному во времени поведению показателей доходов.

	Все данные	Обучающее множество	Проверочное множество
Кол-во	5292	2500	2501
Среднее	0.00037	0.00009	0.00064
Ст. откл.	0.0123	0.0110	0.0138
Асимметрия	-0.0300	-0.2438	0.0562
Эксцесс	21.85	7.47	25.55
$\rho(1)$	0.026	0.093	-0.018
$\rho(2)$	-0.045	0.001	-0.078
$\rho(3)$	-0.030	-0.020	-0.039
$\rho(4)$	0.036	-0.039	0.082
$\rho(5)$	-0.006	0.025	-0.025
Откл. Бартлетта	0.014	0.020	0.020

Таблица 10.1. Ежедневные доходы. Описательная статистика

Все данные целиком не имеют существенной асимметрии, однако обучающее и тестовое множество в отдельности имеют значимую отрицательную и положительную асимметрию, соответственно. Ежедневные доходы имеют эксцесс, намного превышающий эксцесс нормального распределения, и это особенно сильно выражено в последние 10 лет. Обратите внимание также на высокую (но постепенно понижающуюся после 1987 г.) волатильность в тестовом множестве. Любопытно, что обучающее множество сильно положительно (соответственно, отрицательно) скоррелировано с первым (четвертым) лагом,

а тестовое множество сильно положительно (отрицательно) скоррелировано со вторым (четвертым) лагом.

Чтобы минимизировать систематическую ошибку, возникающую при оптимизации, мы ограничились простым перекрестным правилом скользящих средних (CMA = Crossing-Moving-Averages) — это правило торговли пропагандируют Брок и др. [56]. Правило очень простое в том отношении, что в вычислении индикатора не участвуют числа Фибоначчи. Здесь важно, что технический анализ стремится предсказать, главным образом, направление изменения цены (вниз, вверх, на том же уровне), а не величину этого изменения.

По правилу CMA вычисляется совершенно определенная функция $f_{CMA} : \mathbb{R}^L \rightarrow \{-1, 0, 1\}$

$$f_{CMA}(p(t), p(t-1), \dots, p(t-L+1)) = \text{sign} \left(\frac{1}{L} \sum_{j=0}^{L-1} p(t-j) - \frac{1}{L} \sum_{j=0}^{L-1} p(t-j) \right), \quad (1)$$

где L (соответственно, l) — размер долгосрочного (краткосрочного) скользящего среднего. Имеется также полосный вариант правила, содержащий параметр q :

$$f_{CMA}(p(t), p(t-1), \dots, p(t-L+1), q) = \begin{cases} 1, & \text{если } \frac{1}{l} \sum_{j=0}^{l-1} p(t-j) - (1+q) \frac{1}{L} \sum_{j=0}^{L-1} p(t-j) > 0, \\ -1, & \text{если } \frac{1}{l} \sum_{j=0}^{l-1} p(t-j) - (1-q) \frac{1}{L} \sum_{j=0}^{L-1} p(t-j) < 0, \\ 0 & \text{в противном случае.} \end{cases} \quad (2)$$

С точки зрения формализма нейронных сетей это правило можно рассматривать как сеть с заданными весами, которая принимает во входной слой последние L значений цены, имеет два линейных элемента в скрытом слое, которые вычисляют скользящие средние, и один пороговый выходной элемент, выдающий указание на торговлю.

Заметьте, что в такой формулировке обобщенный вариант правила можно легко реализовать, добавив один пороговый элемент, действующий с подходящими весами на два скользящих средних. Очевидно, правило CMA предполагает, что эти два скользящих средних содержат в себе информацию, на основе которой можно предсказывать направление будущих изменений уровня дохода.

При любом варианте спецификации нейронной модели остается вопрос о выборе значений параметров L и l . Следуя, опять-таки, Бро-

ку и др. [56], мы положили L и l равными, соответственно, 50 и 1. Затем мы исследовали чувствительность правила СМА к возрастанию значений q (в интервале от 0 до 0.5). Как видно из табл. 10.2, дневной доход по распоряжениям на покупку (продажу) был положительным (отрицательным), но при этом несильно отличался от безусловного среднего дохода. Более того, доля правильных прогнозов, основанных на сигналах обоих видов («купить > 0 » и «продать < 0 »), была ниже 50%. Однако даже при таких плохих показателях общий доход оказывался положительным из-за асимметрии данных. Любопытно, что 1-50-правило на обучающем множестве работает довольно хорошо, а на новых данных — очень плохо. Это обстоятельство заставляет серьезно усомниться в том, что правило СМА годится для предсказания дохода по акциям Юнилевер. Поскольку два множества обрабатывались совершенно независимо и не содержали перекрывающихся кусков, число распоряжений о сделках всегда было равно размеру множества минус объем долгосрочного скользящего среднего при $q = 0$.

Кроме того, мы пробовали менять значения L и l , оставляя нулевое значение q . В отличие от того, что получили Брок и др. [56], ни один из рядов, представленных в табл. 10.3, не дает существенных результатов.

Если данные автокоррелированы, то, очевидно, путем перемешивания рядов показателей доходов можно устранить последовательную корреляцию. Так как правило СМА на перемешанных рядах давало даже худшие результаты, мы предположили, что исходные (неперемешанные) данные в какой-то степени автокоррелированы. Правило СМА выдавало слишком мало правильных сигналов для того, чтобы его можно было считать надежной моделью торговли, даже при том, что у бутстреп-метода результаты были еще хуже. Вероятно, это означает, что ни одно из использованных скользящих средних не ухватывает существенные черты временного ряда для акций, которые мы выбрали для рассмотрения (или же дело просто в том, что трейдеры акций Юнилевер не пользуются этим временным лагом). Мы не стремились исследовать всевозможные варианты с целью найти подходящий лаг и не пытались применять другие технические правила, например, правило превышения предела изменений цены (TRB = Trading Range Break). Наша цель состояла в том, чтобы выяснить, насколько хорошо нейронные сети могут строить подходящую аппроксимирующую функцию для выявления свойств неэффективности рынка.

Период	Правило	Кол-во «купить»	Кол-во «продать»	Купить	Продать	Купить- продать	Купить > 0	Продать < 0	Успех
Обуч.	(1,50,0)	1313	1137	0.00052 (1.150)	-0.00039 (-1.205)	0.00046 (2.031)	46.0%	45.3%	45.7%
Провер.	(1,50,0)	1674	777	0.00074 (0.221)	0.00047 (-0.303)	0.00035 (0.447)	48.6%	41.7%	46.4%
Обуч.	(1,50,0.02)	783	656	0.00052 (0.964)	-0.00074 (-1.714)	0.00062 (2.167)	44.6%	47.7%	46.0%
Провер.	(1,50,0.02)	1233	454	0.00073 (0.184)	0.00041 (-0.331)	0.00042 (0.424)	48.5%	42.5%	46.9%
Обуч.	(1,50,0.05)	262	285	0.00169 (2.251)	-0.00076 (-1.229)	0.00120 (2.605)	48.5%	49.5%	49.0%
Провер.	(1,50,0.05)	577	191	0.00100 (0.573)	0.00086 (0.212)	0.00054 (0.126)	48.4%	46.1%	47.8%

В столбце 1 указано исследуемое множество.

В столбце 2 указано исследуемое правило.

Первые два числа — длина, соответственно, краткосрочного и долгосрочного скользящих средних, а третье — величина параметра q .

В столбцах 3, 4 приведено количество выданных сигналов, соответственно, на покупку и на продажу.

В столбцах 5, 6 приведены средние показатели дохода по сигналам на покупку и на продажу вместе с соответствующими t -значениями.

Столбец 7 — разности между средними дневными доходами от покупки и продажи и соответствующие t -значения.

В столбцах 8, 9, 10 показаны доли правильных сигналов на покупку, продажу и покупку-или-продажу.

Таблица 10.2. Правило СМА $L - l$ при $L = 50$ и $l = 1$

Период	Правило	Кол-во «купить»	Кол-во «продать»	Купить	Продать	Купить-продать	Купить > 0	Продать < 0	Успех
Обуч.	(50,150,0)	1256	1094	0.00011 (0.074)	0.00027 (0.465)	-0.00007 (-0.346)	45.5%	44.4%	45.0%
Провер.	(50,150,0)	1847	504	0.00051 (-0.312)	0.00114 (0.736)	0.00015 (-0.905)	49.3%	42.7%	47.9%
Обуч.	(1,25,0)	1304	1170	0.00036 (0.724)	-0.00025 (-0.874)	0.00031 (1.383)	45.7%	45.5%	45.6%
Провер.	(1,25,0)	1576	899	0.00096 (0.722)	0.00006 (-1.071)	0.00059 (1.552)	49.7%	44.3%	47.7%
Обуч.	(2,25,0)	1307	1168	0.00033 (0.659)	-0.00023 (-0.815)	0.00028 (1.276)	45.8%	45.7%	45.8%
Провер.	(2,25,0)	1577	899	0.00094 (0.688)	0.00008 (-1.035)	0.00057 (1.493)	49.8%	44.7%	47.9%
Обуч.	(10,50,0)	1298	1152	0.00041 (0.862)	-0.00025 (-0.874)	0.00034 (1.498)	45.5%	45.2%	45.4%
Провер.	(10,50,0)	1670	780	0.00056 (-0.181)	0.00085 (0.373)	0.00011 (-0.485)	48.4%	41.8%	46.3%

Таблица 10.3. Правило SMA $L - l$ при $q=0$

ВОСПРОИЗВЕДЕНИЕ ПРАВИЛА SMA НЕЙРОННОЙ СЕТЬЮ

Статическую нейронно-сетевую модель можно рассматривать как обобщение сети, изображенной на рис. 10.1, и поэтому она похожа на правило SMA с той разницей, что значения величин L , l и параметра q заранее не фиксируются. Нейронная сеть должна выделять существенные показатели и оценивать эффективные значения. В соответствии с принципами технического анализа, требуется, глядя на временной ряд цен, выдавать сигнал на покупку, продажу или удержание позиции при каждом поступлении новых данных. При различной ширине временного окна и числе показателей принятые решения будут различными.

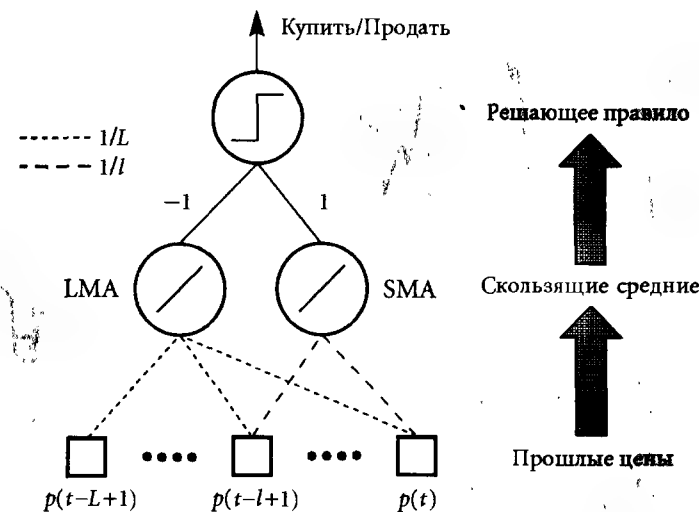


Рис. 10.1. Нейронная сеть, реализующая правило SMA

Чтобы уменьшить число возможных моделей, мы ограничились размером временного окна для прошлых цен в пять последних рабочих дней (1 неделя). Такой выбор в чем-то произволен, но он, во всяком случае, гарантирует, что даже при большом числе скрытых элементов мы будем иметь достаточный объем данных (2500 точек на обучение) по отношению к количеству весов (см. [35]). Каждая запись, состоящая из пяти последовательных цен, была перемасштабирована так, чтобы значения лежали от 0 до 1. При этом истинные значения утрачиваются, и все входные записи вкладываются в гиперкуб $[0,1]^5$. Таким образом, при любом уровне цен гарантируется инвариантность преобразования входной записи. Если имеются две записи i и j , причем цены в j равны соответствующим ценам в $i + f + 10$ флори-

нов, то структура обеих записей будет одинакова. Такое перемасштабирование не лишено смысла, так как трейдер-человек обычно оценивает данные временного ряда в относительном выражении с помощью стандартных приемов. Результат описанного преобразования показан на рис. 10.2.

Целевой переменной является направление завтрашнего сдвига цены, или знак дохода. Выходной сигнал кодируется следующим образом:

- 0 — распоряжение о продаже,
- 0.5 — нейтральный сигнал,
- 1 — распоряжение о покупке.

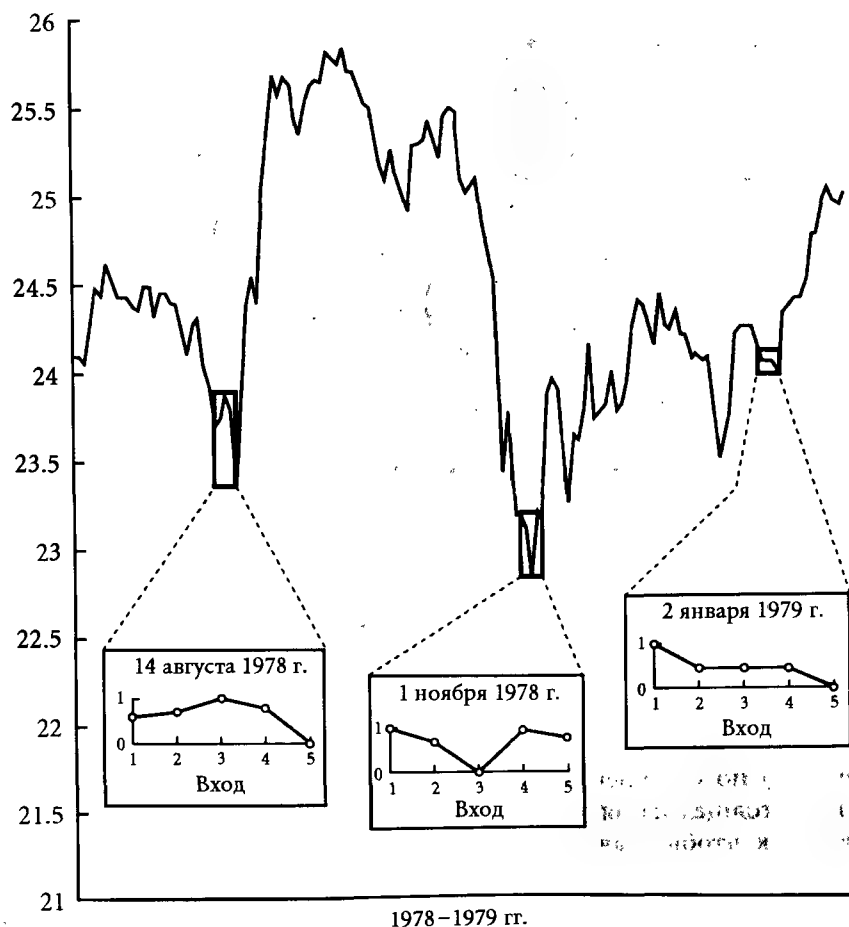


Рис. 10.2. Результаты перемасштабирования различных входных сигналов

Как уже говорилось, правило СМА дает относительно лучшие результаты на неперемешанных данных. Тем самым мы неявно принимаем, что имеет место некоторая неэффективность (и, следовательно, предсказуемость) рынка. На это указывают небольшие отклонения от гиперплоскости $s = 0.5$, которые можно видеть на рис. 10.3. При эффективном рынке вероятность выдачи сигнала на покупку (или продажу) на основе информации, относящейся к прошлому, равна 50% (0.5). В такой ситуации решающая поверхность в пространстве задержек (входном пространстве сети) должна в точности совпадать с гиперплоскостью $s = 0.5$. В нашем случае определенная степень неэффективности приводит к тому, что некоторые области в пространстве входов соответствуют значениям, немного отличным от 0.5, так что решающая поверхность имеет небольшие локальные отклонения от гиперплоскости $s = 0.5$.

Сигнал ($t+1$)

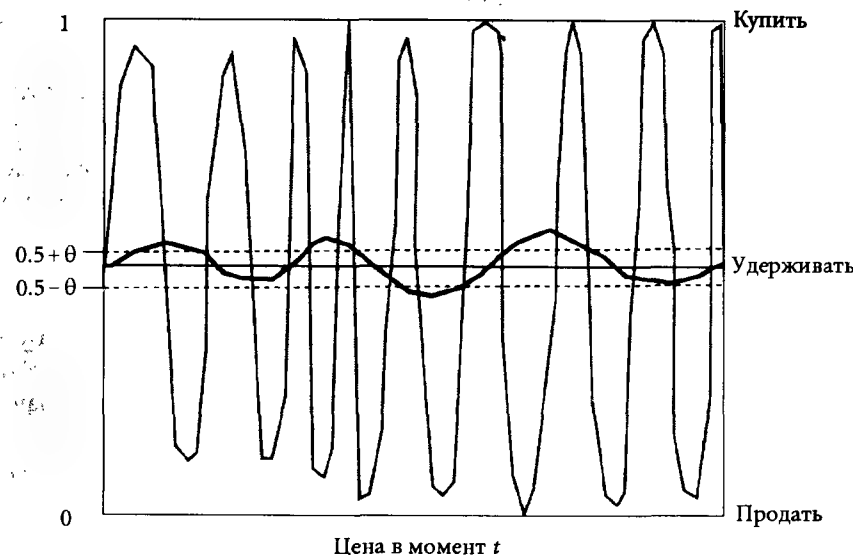


Рис. 10.3. Неэффективность — наличие отклонений от гиперплоскости $s = 0.5$

Нейронные сети хорошо приспособлены для отображения таких поверхностей в случаях, когда данные позволяют использовать достаточное количество нелинейных средств (скрытых элементов). Возникает следующая проблема: коль скоро мы определили неэффективность как малые отклонения от 0.5-гиперплоскости, нам, очевидно, не нужно, чтобы сеть чересчур точно приближала обучающее множество (светлая волнистая линия). Конкретнее, нам нужно найти такое решение в фазовом пространстве модели, которое бы несильно отходило

от 0.5-гиперплоскости (жирная кривая), или, иначе говоря, такое решение, которое дало бы квадратный корень из среднеквадратичной ошибки RMSE, близкий к 0.5 (за вычетом нейтральных сигналов). Степень обучения трудно отследить с помощью подтверждающего множества, так как при исходных малых случайных весах (сплошная 0.5-линия) ошибка RMSE уже будет близка к 0.5. Мы не применяли известный метод перекрестного подтверждения, потому что процедура обучения обычно прекращалась сразу же, когда начинал работать обучающий алгоритм (в зависимости от порядка подачи образцов обучающего множества). Вместо этого мы использовали методы регуляризации.

Для того чтобы сеть не выдавала слишком сложную решающую поверхность, в критерий ошибки был добавлен штрафной член и соответствующим образом изменен алгоритм обучения. В результате мы минимизировали следующее выражение:

$$E = \frac{1}{2} \sum_{k=1}^P (d_k - s_k)^2 + \frac{\eta}{2} \sum_{i \rightarrow j} W_{ij}^2, \quad (3)$$

где d_k — правильный сигнал для k -го образца, а s_k — сигнал, который выдает сеть. В формулу включен параметр η для того, чтобы веса не могли принимать очень больших значений, — это сужало бы класс функций, которые сеть способна реализовать. Такой критерий ошибки уже обсуждался в гл. 1 (соотношение (17)).

Далее, был введен параметр θ , чтобы уменьшить число сигналов о покупке/продаже, выдаваемых сетью. Смысл этого очевиден: количество сделок в течение одного дня ограничено из-за расходов на совершение сделок (транзакционных издержек). Сеть никогда не даст на выходе значение, в точности равное 0.5, и мы относили к нейтральным все сигналы, лежащие от 0.5 на расстоянии плюс/минус θ .

Для оценки качества работы сети использовались три показателя, аналогичные тем, которые были взяты в табл. 10.3:

- $R(\text{купить})$ — средний доход по сигналам на покупку,
- $R(\text{продать})$ — средний доход по сигналам на продажу,
- $R(\text{купить/продать})$ — средний доход по сигналам обоих видов.

РЕЗУЛЬТАТЫ РАБОТЫ НЕЙРОННОЙ СЕТИ

5-20-1-сеть обучалась в течение 3000 эпох. Число скрытых элементов было выбрано волевым образом из расчета, чтобы было достаточно вычислительных ресурсов для размещения сигмоидов в гиперпространстве. Коэффициент обучения ε был взят равным 0.25, а ограничивающий параметр η — равным 0.0005. Так как обучение прекраща-

ется после фиксированного числа эпох и нет оснований ожидать, что сходимость будет достигнута, то и не следует обращать слишком большое внимание на числовые значения параметров. Однако на моделях с нужными свойствами даже ранняя остановка обучения (как это имеет место в нашем случае) может дать определенный результат. При этом произвол в выборе параметров является главным недостатком нейронно-сетевого подхода.

В табл. 10.4 представлены результаты, которые показала обученная сеть на новых данных при различных значениях отсекающего порога. Транзакционные издержки, по-прежнему, в расчет не брались. Интересно, что с повышением порога (т.е. при уменьшении количества сделок) возрастали доли правильных сигналов и на покупку, и на продажу. По-видимому, те неэффективности (т.е. малые отклонения от 0.5-линии на рис. 10.3), которые имели место в периоде обучения (1970-е годы), подтвердились в проверочном периоде (1980-е годы).

При $\theta = 0$ доля правильных сигналов была менее 50%, потому что в этом случае каждый раз должен был выдаваться сигнал (на покупку или продажу), тогда как в значениях целевой переменной до 10% могут составлять распоряжения удерживать имеющуюся позицию. На рис. 10.4 показана зависимость достигнутого дохода от пороговых значений. Средний доход на одну сделку (покупку или продажу) примерно 0.13% выигрывает в сравнении с безусловным средним доходом 0.064% (для наибольших значений θ).

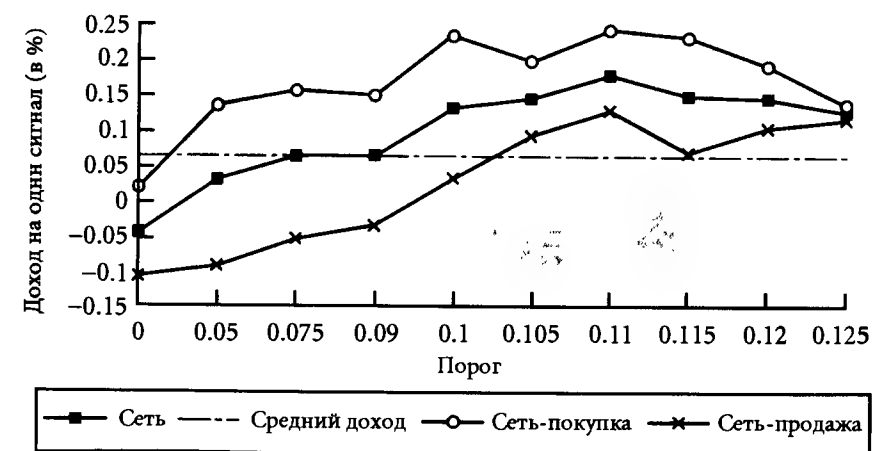


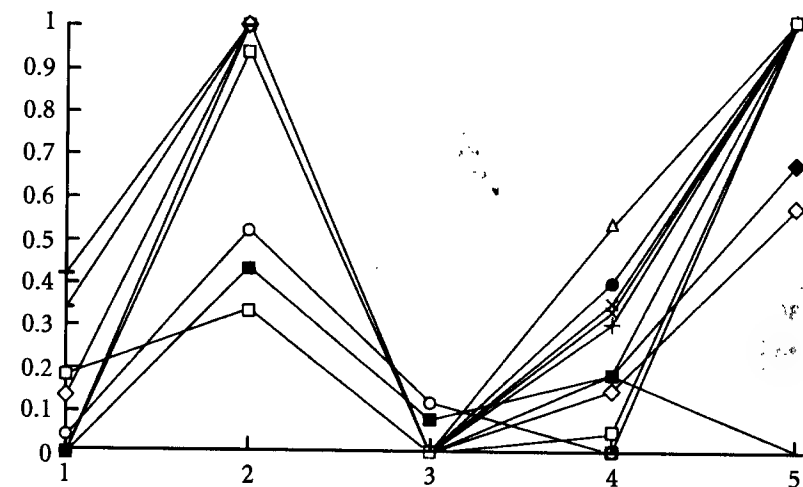
Рис. 10.4. Совокупный доход для возрастающих значений θ

Модель торговли достигает наилучших показателей при количестве сделок около 250 в расчете на все тестовое множество (или около 25 сделок в год). Дальнейшее уменьшение числа сделок путем поднятия порога приводит к уменьшению дохода на одну сделку.

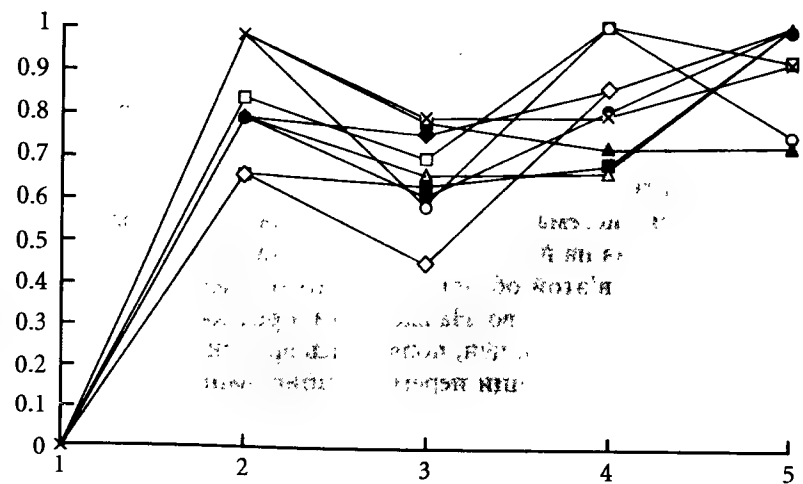
Тема	Кол-во «купить»	Кол-во «продать»	Купить	Продать	Купить- продать	Купить > 0	Продать < 0	Успех
0	1227	1273	0.0002024 (-0.907)	0.0010608 (0.885)	-0.0008584 (-1.551)	49.71%	43.91%	46.76%
0.05	579	505	0.0013548 (1.121)	0.0009169 (0.411)	0.0004379 (0.520)	52.33%	43.56%	48.25%
0.075	352	280	0.0015585 (1.167)	0.0005341 (-0.121)	0.0010244 (0.925)	54.26%	43.93%	49.68%
0.09	237	211	0.0014866 (0.901)	0.0003691 (-0.273)	0.0011175 (0.854)	53.16%	45.97%	49.78%
0.1	167	172	0.0023428 (1.541)	-0.0002775 (-0.841)	0.0026203 (1.744)	56.29%	49.42%	52.80%
0.105	143	155	0.0019836 (1.130)	-0.0008902 (-1.336)	0.0028738 (1.792)	54.55%	50.97%	52.68%
0.11	125	142	0.0024005 (1.389)	-0.0012447 (-1.579)	0.0036452 (2.149)	56.00%	50.70%	53.18%
0.115	113	124	0.0023238 (1.266)	-0.0006691 (-1.028)	0.0029929 (1.664)	53.98%	49.19%	51.48%
0.12	93	113	0.0019135 (0.872)	-0.0009954 (-1.229)	0.0029089 (1.502)	54.84%	50.44%	52.43%
0.125	79	106	0.0013371 (0.441)	-0.0011062 (-1.273)	0.0024432 (1.188)	56.96%	51.89%	54.05%

Таблица 10.4. Результаты на новых данных

Чтобы лучше представить себе механизмы, стоящие за сигналами на покупку или продажу, мы изобразили на графике (см. рис. 10.5) восемь входных образцов, породивших правильные сигналы, соответственно, на покупку и продажу. Мы не исследовали каждое значимое отклонение от 0.5-линии по отдельности (это можно сделать, например, при помощи техники кластеризации весов состояния методом Гормана и Сейновского [127], который обсуждался в гл. 4), и поэтому из результатов, относящихся к этим восьми примерам, нельзя делать обобщения. Однако, образцы внутри каждой группы обладают определенными общими чертами, и это говорит о том, что они соответствуют одним и тем же областям в пространстве входов.



Образцы входных данных с большой положительной активацией выхода



Образцы входных данных с большой отрицательной активацией выхода

Рис. 10.5. Примеры активации

Наконец, есть еще один способ оценить качество модели: подсчитать, сколько денег вы заработаете с ее помощью. На рис. 10.6 представлены соответствующие результаты для $\theta = 0.15$. Сигналы на покупку или продажу дают возможность трейдеру свести свой дневной баланс с плюсом. Ложные же сигналы уменьшают суммарный доход. Нейтральные сигналы игнорируются, т.е. полученный доход переходит на следующий день. При торговле по данной системе, в предположении, что все сигналы были правильные, будет получено 14% от максимально достижимого дохода. Хорошо видно, что система дает отклонение в положительную сторону от линии случайного прогноза. Однако, если бы учитывались транзакционные издержки, то прибыли могло и не быть.

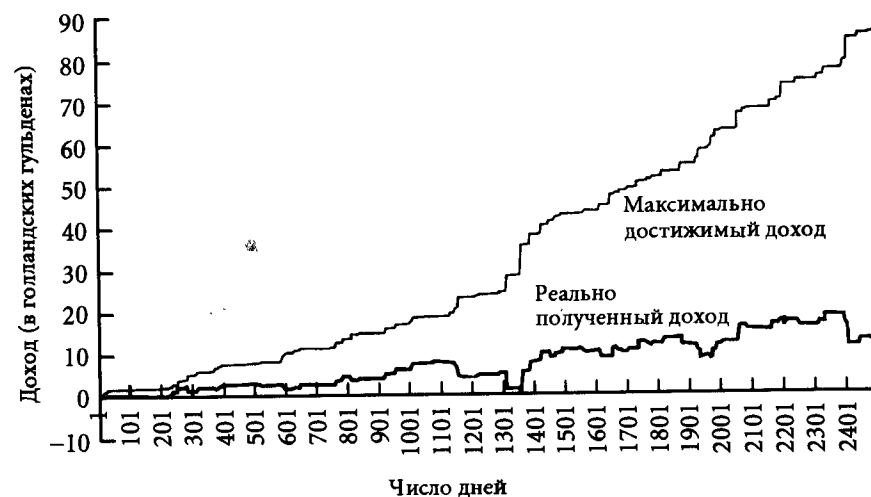


Рис. 10.6. Нереализованная прибыль

ОБСУЖДЕНИЕ

Применение нейронных сетей в этой задаче показало, что статическая нелинейная система может быть обучена так, чтобы выполнять технический анализ на начальном уровне. Наш подход отличается от других, известных в этой области, тем, что мы не пытаемся оптимизировать входное множество. На вход сети, предсказывающей направление будущих изменений цен, подавались временные ряды данных в исходном виде. При помощи перемасштабирования данных достигалась инвариантность пространства входов. Затем сеть обучалась с использованием регуляризирующего штрафного члена, ограничивающего положение интерполирующей поверхности, построенной нейронной сетью. Модель выдавала до 54% правильных сигналов.

Этот результат, по-видимому, свидетельствует о присутствии свойств неэффективности рынка (которые определяются как малые отклонения от 0.5-гиперплоскости) и о том, что нейронная сеть является адекватным инструментом для их обнаружения. Однако необходимы дальнейшие исследования, прежде чем данный метод можно будет использовать непосредственно в торговом зале. При том, что результаты оказались многообещающими, прибыльной стратегии торговли не выработано. Мы хотим высказать ряд предложений, реализация которых, по нашему мнению, могла бы способствовать выработке более совершенной стратегии торговли.

Во-первых, следует расширить базу данных и таким образом повысить представительность данных и увеличить размерность входных образцов. Это можно сделать, включив в рассмотрение акции других компаний, похожих на Юнилевер. При этом остается неясным, как определить степень сходства.

Во-вторых, можно было бы использовать данные котировок с мониторов (тикеров) и таким путем увеличить частоту поступления данных и получить возможность использовать экономический (или операционный) масштаб времени, когда время сжимается (соответственно, растягивается) по сравнению с календарем. Не следует также недооценивать возможность подключения подсказок (см. [1], [2]).

В-третьих, сеть, использующая радиальные функции, позволяет получить локально более точные отображения, чем классическая сигмоидальная, и за счет этого, по-видимому, можно добиться более точного распознавания «неэффективностей» в пространстве входов. Далее, доверительные интервалы (полосы ошибок) можно вычислять по методам, которые были предложены МакКеем [184] и Ле Каном [174]. Правда, эти методы предполагают сходимость обучающего алгоритма.

Четвертое: улучшение характеристик может быть получено за счет использования другого критерия ошибок, который включал бы в себя транзакционные издержки и, наряду с абсолютным значением, раздельный подсчет потоков платежей в обоих направлениях. Другая возможность заключается в том, чтобы ввести дополнительный выходной элемент, выдающий абсолютное значение дохода.

И последнее: модель может быть включена в качестве составной части в многосетевую среду принятия решений, а полученная общая производительность — измеряться, исходя из заданного решающего правила (см. [290]). Наконец, более динамичные подходы можно получить, используя рекурсивные сети с механизмами обратной связи.

Список литературы

1. Abu-Mostafa, Y. S. (1993 a). 'A method for learning from hints'. In: Hanson, S., Cowan, J., and Giles, C. (eds), *Advances in Neural Information Processing Systems*. Vol. 5, 73–80
2. Abu-Mostafa, Y. S. (1993b). Hints and the VC dimension. *Neural Computation*, 5, 278–288
3. Abu-Mostafa, Y. S. (1994). 'Financial market applications of learning from hints', to appear in Refenes, A. N. (ed), *Neural Networks in the Capital Markets*, London: Wiley
4. Adler, M. and Dumas, D. (1983) 'International portfolio selection and corporate finance: A synthesis', *Journal of Finance*, 38:3, 925–984
5. Ahmadi, H. (1993), 'Testability of the arbitrage pricing theory by neural networks'. In: Trippi, R. R. and Turban, E. (eds), *Neural Networks in Finance & Investment*, Chicago: Probus Publishing Company, 421–432
6. Aleksander, I. and Morton, H. (1990), *An Introduction to Neural Computing*, London: Chapman & Hall
7. Altman, E. I. (1968), 'Financial ratios discriminant analysis and the prediction of failure', *Journal of Finance*, Vol. 23, No. 4
8. Altman, E. I. (1983), *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*, NY: John Wiley & Sons
9. Altman, E. I. and Brenner, M. (1981), 'Information effects and stock market response to signs of firm deterioration', *Journal of Financial and Quantitative Analysis*, Vol. XVI. No. 1, 39–51
10. Altman, E. I. and McGough, T. P. (1974), 'Evaluation of a company as going concern', *Journal of Accountancy*, 138:6, 50–57
11. Altman, E. I., Avery, R. B., Eisenbeis, R. A. and Sinkey, J. F. (1981), *Application of Classification Techniques in Business, Banking and Finance*, Greenwich, Connecticut: JAI Press
12. Altman, E. I., Haldeman, R. G. and Narayanan, P. (1977), 'ZETA Analysis: A new model to identify bankruptcy risk of corporations', *Journal of Banking and Finance*, 1:1, 29–54
13. Anderson, R. W. (1993), 'Biased random-walk learning: A neurobiological correlate to trial-and-error', to appear in *Progress in Neural Networks*
14. Anthony, J. H. (1988), 'The interrelation of stock and options market trading-volume data', *Journal of Finance*, 43, 949–964
15. Argenti, J. (1976), *Corporate Collapse*, England: McGraw-Hill
16. Arthur, W. B. (1990), 'Positive feedbacks in the economy', *Scientific American*, Feb. 80–85
17. Ash, T. (1989), 'Dynamic node creation in backpropagation networks', *Connection Science*, Vol. 1, 4
18. Auten, G. E. and Robb, E. H. (1976), 'A general model for state tax revenue analysis', *National Tax Journal* 29, 422–447
19. Baestaens, D.-E. (1994), 'Distributional analysis to model atypical behaviour', to appear in *European Journal of Operational Research*, 4
20. Baestaens, D.-E. and Van den Bergh, W.-M. (eds) (1993a), *Proceedings of 16th FinBel-Dag Seminar on 'Intelligent Structuring in Chaos'*, Erasmus University Rotterdam, Department of Finance, 27 May
21. Baestaens, D.-E., Van den Bergh, W.-M. and Vaudrey, H. (1993b), 'Qualitative credit assessment using a neural classifier', *Neural Network World*, 3:6, 605–624
22. Baestaens, D.-E., Van den Bergh, W.-M. and Vaudrey, H. (1993c), 'Estimating tax inflows at a public institution'. In: Refenes, A. N., *NnCM'93: Proceedings of the First International Workshop on Neural Networks in the Capital Markets*, London Business School, 18–19 Nov.
23. Baestaens, D.-E. and Van den Bergh, W.-M. (1992), 'Tracking the Amsterdam stock index using Hecht-Nielsen's knowledge net', *Neural Network World*, 2, 543–651
24. Baghestani, H. and McNown, R. (1992), 'Forecasting the Federal budget with time-series models', *Journal of Forecasting*, Vol. 11, p. 127–139
25. Bailey, D. L., Thompson, D. M. and Feinstein, J. L. (1988), 'Options trading using neural networks', *Proceedings Neuro-Nimes*, 15–17 Nov. 395–402
26. Banking Law, (1992), *The Banking Law of January 31, 1989, incl. Act of December 19*
27. Barnard, J. R. and Dent, W. T. (1979), 'State tax revenues — New methods of forecasting', *The Annals of Regional Science* 13, 1–15
28. Barndorff-Nielsen, O.-E., Jensen, J. L. and Kendall, W. S. (eds) (1993), *Networks and Chaos: Statistical and Probabilistic Aspects*, London: Chapman & Hall

29. Barnes, P. (1987), 'The analysis and use of financial ratios: A review article', *Journal of Business Finance & Accounting*, Winter, 14 (4), 449-462
30. Barron, A. (1984), 'Predicted squared error: a criterion for automatic model selection'. In: Farlow, S. (ed), *Self-Organizing Methods in Modeling*, NY: Marcel Dekker
31. Baum, E. B. and Haussler, D. (1989), 'What size net gives valid generalization?', *Neural Computation*, 1, 151-160
32. Beaver, W. H. (1966), 'Financial ratios as predictors of failure', *Journal of Accounting Research*, Selected studies, Supplement to 5, 71-111
33. Beaver, W. H. (1968), 'Market prices, financial ratios and prediction of failure', *Journal of Accounting Research*, Autumn
34. Beckers, S. (1981), 'Standard deviations implied in option prices as predictors of future stock price volatility', *Journal of Banking and Finance*, 363-381
35. Bellman, R. E. (1961), *Adaptive Control Processes*, Princeton University Press
36. Beltratti, A. and Margarita, S. 'Simulating an artificial adaptive stock market', Università di Torino and Fondazione ENI Enrico Mattei, 1-27
37. Benninga, S. (1989), *Numerical Techniques in Finance*, MA: MIT Press
38. Berger, A. N. and Udell, G. F. (1990), 'Collateral, loan quality, and bank risk', *Journal of Monetary Economics*, No. 25, 21-42
39. Berkman, H. (1992), 'Trading systems and liquidity on securities markets, a study of the European Options Exchange', Ph. D. Dissertation, Erasmus University, Rotterdam
40. Bernasconi, J. (1991), 'Learning strategies for neurocontrollers'. In: Würtz, D. and Murtagh, F. (eds), *Proceedings International Workshop on Parallel Problem Solving From Nature: Applications in Statistics and Economics PASE-91*, Zürich, 38-44
41. Bhattacharya, M. (1987), 'Price changes of related securities: The case of options and stocks', *Journal of Financial and Quantitative Analysis*, 22, 1-15
42. Black, F. (1975), 'Facts and fantasy in the use of options', *Financial Analysts Journal*, 31, 36-41
43. Black, F. and Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy*, 637-659

44. Blum, A. (1992), *Neural Networks in C++: An Object-Oriented Framework for Building Connectionist Systems*, NY: John Wiley & Sons
45. Boddā, R. E. J. (1986), *Opties: De Winst van Rationeel Beleggen*, Utrecht: Veen uitgevers
46. Bollerslev, T, Chou, R. Y. and Kroner, K. F. (1992), 'ARCH modeling in finance', *Journal of Econometrics*, 5-59
47. Bosarge, Jr. W. E. (1993), 'Adaptive processes to exploit the nonlinear structure of financial markets'. In: Trippi, R. R. and Turban, E. (eds), *Neural Networks in Finance & Investment*, Chicago: Probus Publishing Company, 371-402
48. Box, G. E. P. and Jenkins, G. M., (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day
49. Bramson, M. J. and Hoptroff, R. G. (1990), 'Forecasting the economic cycle: a neural network approach'. In: Murtagh, F. (ed), *Proceedings PASE 1990, Neural Networks for Statistical and Economic Data*, 121-153
50. Brent, R. P. (1991), 'Fast training algorithms for multilayer neural nets', *IEEE Transactions on Neural Networks*, Vol. 2, 346-354
51. Brock, W. A. (1986), 'Distinguishing random and deterministic systems: abridged version', *Journal of Economic Theory*, 40, 168-195
52. Brock, W. A. and Dechert, W. D., (1989) 'Statistical inference theory for measures of complexity in Chaos Theory and Nonlinear Science'. In: Abraham, N. B. et al. (eds), *Measures of Complexity and Chaos*, New York: Plenum Press, 79-97
53. Brock, W. A. and Baek, E. G. (1991), 'The theory of statistical inference for nonlinear science: Gauge functions, complexity measures, and instability measures', *RES*, Jul. 1-61
54. Brock, W. A. and Chamberlain, G. (1984), 'Spectral analysis cannot tell a macro-econometrician whether his time series came from a stochastic or a deterministic economy', *Social Systems Research Institute, University of Wisconsin*, Oct. 1-11
55. Brock, W. A., Hsieh, D. A. and LeBaron, B. (1991), *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*, Cambridge: MIT Press
56. Brock, W. A., Lakonishok, J. and LeBaron, B. (1992), 'Simple technical trading rules and the stochastic properties of stock returns', *Journal of Finance*, XLVII: 5, 1731-1764
57. (1992), *FKSEC, A Macro-econometric Model for the Netherlands*, Stenfert Kroese, Leiden

58. Carling, A. (1992), *Introducing Neural Networks*, Cheshire, UK: Sigma Press
59. Casdagli, M. and Eubank, S. (1992), *Nonlinear Modelling and Forecasting*, California: Addison-Wesley
60. Casey, C. and Bartczak, N. (1985), 'Using operating cash flow data to predict financial distress: Some extensions', *Journal of Accounting Research*, Vol. 23, No. 1, 384-401
61. Caudill, Maureen (1989), 'Neural networks primes', *AI Expert*, 1-63
62. Chan, K. C. (1988), 'On the contrarian investment strategy', *Journal of Business*, April, 147-163
63. Chan, K. C. and Lakonishok, J. (1993), 'Are the reports of beta's death premature?', *Journal of Portfolio Management*, Summer, 51-62
64. Chauvin, Y. (1989), 'A back-propagation algorithm with optimal use of hidden units'. In: Touretzky, D. S. (ed), *Advances in Neural Information Processing Systems*, San Mateo, California: Kaufman Publishing
65. Chauvin, Y. (1991), 'Investment decision learning using parallel networks: application to currency trading', in *Proceedings PASE 1991*, Zürich, Dec. Parallel problem solving-applications in statistics and economics, 137W6
66. Chen, N. F. (1983), 'Some empirical tests of the theory of arbitrage pricing', *Journal of Finance*, 38, 393-414
67. Chen, N. F., Roll, R. and Ross, S. A. (1986), 'Economic forces and the stock market', *Journal of Business*, 59, 383-403
68. Coats, P. K., (1988), 'Why expert systems fail?', *Financial Management*, Autumn, 77-86
69. Colin, A. (1991a), 'Exchange rate forecasting at Citibank London', *Proceedings of Neural Computing*, London
70. Colin, A. (1991b), 'Machine learning techniques for foreign exchange trading'. In: *Proceedings PASE 1991*, Zürich, Dec. Parallel problem solving-applications in statistics and economics, 115
71. Collard, J. E. (1991a), 'A B-P ANN commodity trader', Martingale Research Corporation, Texas
72. Collard, J. E. (1991b), 'Commodity trading with a two year old', Martingale Research Corporation, Texas
73. Collins, R. A. and Green, R. D. (1982), 'Statistical methods for bankruptcy forecasting', *Journal of Economics and Business*, 32, 349-354

74. Connor, J. and Atlas, L. (1991), 'Recurrent neural networks and time series prediction', *Proceedings IJCNN 1*, 301-306
75. Copeland, T. E. and Weston, J. F. (1988), *Financial Theory and Corporate Policy*, Massachusetts: Addison-Wesley
76. Cover, T. M. (1965), 'Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition', *IEEE Transactions on Electronic Computers*, 14, 326-334
77. Cowan, J. D. (1990), 'Neural networks: the early days'. In: Touretzky, D. S. (ed), *Advances in Neural Information Processing Systems*, San Mateo, California: Kaufman Publishing, 828-842
78. Crick, F. (1989), 'The recent excitement about neural networks', *Nature*, 337, 129-132
79. Cybenko, G. (1988), 'Approximations by superpositions of a sigmoidal function', Techn. Rep. No. 856, Urbana Univ. of Illinois
80. Deboeck, G. J. (ed) (1994), *The Trading Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, NY: John Wiley & Sons
81. De Bondt, W. F. M. (1991), 'What do economists know about the stock market?', *Journal of Portfolio Management*, Winter, 84-91
82. De Gooijer, J. G. (1989), 'Testing non-linearities in world stock market prices', *Economics Letters*, 31, 31-35
83. De Groot, C. and Würtz, D. (1991), 'Time series analysis with connectionist networks', *Proceedings PASE 1991*, Zürich, Dec. Parallel problem solving-applications in statistics and economics, 87-96
84. De Groot, C. (1993), 'Nonlinear time series analysis with connectionist networks', IPS Research Report No. 93-03, Diss. ETH No. 10038, Zürich
85. De Groot, C. and Würtz, D. (1992), 'Plain backpropagation and advanced optimization algorithms: a comparative study', Working paper, Interdisziplinäres Projektzentrum für Supercomputing, ETH-Zentrum and Institut für Theoretische Physik, ETH-Honggerberg, Zürich, 1-10
86. De Grauwe, P. and Decupere, D. (1992), 'Psychological barriers in the foreign exchange market', *CEPR Working Paper*
87. Denoeux, T. and Lengell, R. (1993), 'Initializing back propagation networks with prototypes', *Neural Networks*, Vol. 6, 351-363
88. De Sieno, D. (1988), 'Adding a conscience to competitive learning'. In: *Proceedings of the IEEE International Conference on Neural Networks*, Vol. 1

89. Dijk, Vi. W. van (1993), 'Neural networks and data analysis', a marketing application, EUR, Dep. of computer science, Mar. 1-109
90. Ding, Z., Granger, C. W. J. and Engle, R. F. (1993), 'A long memory property of stock market returns and a new model', *Journal of Empirical Finance*, 1:1, 83-106
91. Doherty, C. (1990), 'Recurrent cascade-correlation architecture vs. the Box & Jenkins method on forecasting univariate time series'. In: Murtagh, F. (ed), *Proceedings PASE 1990*, Neural networks for statistical and economic data, 167-177
92. Dorizzi, B., Duval, J. M. and Debar H. (1992), 'Utilisation de reseaux recurrences pour la prevision de consommation electrique', *Proceedings of NeuroNimes 1992*
93. Dutta, S. and Shekhar, S. (1988), 'Bond rating: A non-conservative application of neural networks', *Proceedings of the IEEE International Conference on Neural Networks*, II, 443-450
94. Eaton, M. and Collins, B. J. (1990), 'Neural networks front end to an expertsystem for decision taking in an uncertain environment'. In: Murtagh, F. (ed), *Proceedings PASE 1990*, Neural networks for statistical and economic data, 179-182
95. Economist Intelligence Unit Limited (1993a), 'Poland country profile 1992-1993', Economist Intelligence Unit, London, No. 1
96. Economist Intelligence Unit Limited (1993b), 'Poland country report', Economist Intelligence Unit, London, No. 1
97. Economist, The (1992a), 'Neural networks: The campaign for real neurons', *The Economist*, 9 May, 116
98. Economist, The (1992b), 'Beating the market? Yes, it can be done', *The Economist*, 5 Dec. 25-27
99. Economist, The (1993a), 'Poland's economic reforms: If it works, you've fixed it', *The Economist*, 23 Jan. 21-25
100. Economist, The (1993b), 'The mathematics of markets: A survey of the frontiers of finance', *The Economist*, 9 Oct. 1-20
101. Eisenbeis, R. A. (1977), 'Pitfalls in the application of discriminant analysis in business, finance, and economics', *Journal of Finance*, June, 875-899
102. Errunza, V. R. (1983), 'Emerging markets: new opportunities for improving global portfolio performance', *Financial Analysts Journal*, 39, 51-58
103. Eubank, R. L. (1988), *Spline Smoothing and Non-parametric Regression*, NY: Marcel Dekker Inc

104. Fahlman, S. E. (1988), 'An empirical study of learning speed in back-propagation networks'. Technical Report CMU-CS-88-162, CMU
105. Fahlman, S. E. (1992), Comments on comp.ai.neural-nets, item 2198
106. Fahlman, S. E. and Lebiere, C. (1990), 'The cascade-correlation learning architecture'. In: Touretzky, D. S. (ed), *Advances in Neural Information Processing Systems*, San Mateo, California: Kaufman Publishing
107. Fama, E. F. and Miller, M. H. (1972), *The Theory of Finance*, New York: Holt, Rinehart and Winston
108. Fama, E. F. and Gibbons, M. R. (1984), 'A comparison of inflation forecasts', *Journal of Monetary Economics*, 13, 327-348
109. Fase, M. M. G., Beckers, C. E., Kemna, A. G. Z. and de Wilde, S. (1990), *Tussen Rokin en Damrak, de Wisselwerking Tussen de Effectenen Optiebeurs in Amsterdam*, Amsterdam Stock Exchange
110. Ferson, W. A. and Harvey, C. R. (1991), Amsterdam Stock Exchange 'Sources of predictability in portfolio returns', *Financial Analysts Journal*, May-June, 49-56
111. Fisher, R. A. (1936), 'The use of multiple measurements in taxonomic problems', *Ann. Eugenics*, 7, 179-188
112. Fogarty, T. C. (1991), 'Credit scoring and control applications of the genetic algorithm'. In: *Proceedings PASE 1991*, Zürich Dec. Parallel problem solving-applications in statistics and economics, 147-148
113. Fogelman-Soulie, F. (1992), 'Neural networks: State of the art'. In: *3rd International Workshop on Parallel Applications in Statistics and Economics*, Prague (PASE 92), 7-8 Dec. 1-37
114. Frain, John, (1992), 'Complex dynamics and chaos in economics', Conference on Analysis and Forecasting of Time Series, Torino, June
115. Frane, J. W. (1977), 'A note on checking tolerance in matrix inversion and regression', *Technometrics*, 19, 513-514
116. Frean, M. (1990), 'The Upstart algorithm: a method for constructing and training feedforward neural networks', *Neural Computation*, Vol. 2, 2, 198-209
117. Fullerton, Jr. T. M. (1989), 'A composite approach to forecasting state government revenues: Case study of the Idaho sales tax', *International Journal of Forecasting* 5, 373-380
118. Funahashi, K. I. (1989), 'On the approximate realization of continuous mappings by neural networks', *Neural Networks* 2, 183
119. Gabr, M. M. and Subba Rao, T. (1981), 'The estimation and prediction of subset bilinear time series models with applications', *Journal of Time Series Analysis*, 2, 155

120. Garson, G. D. (1991), 'Interpreting neural networks connection weights', *AI Expert*, 47-51
121. George, A. (1991), 'Qualitative analysis: Evaluating a borrower's management and business risks', *The Journal of Commercial Bank Lending*, Aug. 6-16
122. Ghaziri, H. EL (1991), 'An efficient neural network algorithm for routing problems', in *Proceedings PASE 1991*, Zürich, Dec. Parallel problem solving-applications in statistics and economics, 165
123. Gielen, S. and Kappen, B. (eds) (1993), *Proceedings of the International Conference on Artificial Neural Networks*, ICANN-93, Amsterdam, London: Springer-Verlag
124. Gifi, A. (1990), *Nonlinear Multivariate Analysis*, NY: John Wiley & Sons
125. Gilbert, L. R., Menon, K. and Schwartz, K. B. (1990), 'Predicting bankruptcy for firms in financial distress', *Journal of Business, Finance & Accounting*, No. 17 (1), Spring, 161-171
126. Girosi, F. and Poggio T. (1990), 'Networks and the best approximation property', *Biological Cybernetics*, 63, 169-176
127. Gorman, R. P. and Sejnowski T. J. (1988), 'Analysis of hidden units in a layered network trained to classify sonar targets', *Neural Networks*, Vol. 1, 75-89
128. Graddy, D. B. and Spencer, A. H. (1990), *Managing Commercial Banks: Community, Regional and Global*, Englewood Cliffs, N. J.: Prentice Hall
129. Granger, C. W. J. and Anderson, T. W. (1978), *Introduction to Bilinear Time Series Models*, Gottingen: Vandenhoeck und Ruprecht
130. Hakala, J., Goerke, N. and Fahner, G. (1991), 'HENAMnet: an alternative neural net approach for prediction of chaotic time series', in *Proceedings PASE 1991*, Zürich, Dec. Parallel problem solving-applications in statistics and economics, 153-155
131. Hampshire, J. B. and Pearlmutter, B. A. (1990), 'Equivalence proof for multilayer perceptron classifiers and the Bayesian discriminant function', *Proceedings of the 1990 Connectionist Models Summer School*
132. Hanson, S. J. and Pratt, L. (1989), 'A comparison of different biases for minimal network construction with back-propagation'. In: Touretzky, D. S. (ed), *Advances in Neural Information Processing Systems*, San Mateo, California: Kaufman Publishing
133. Hansson, P. A. (1991), 'Chaos: implications for forecasting', *Futures*, No. 1, 50-58

134. Harp, S. A., Samad, T. and Guha, A. (1989), 'Design application-specific neural networks using the genetic algorithm'. In: Touretzky, D. S. (ed), *Advances in Neural Information Processing Systems*, San Mateo, California: Kaufman Publishing, 447-454
135. Hart, A. (1992), 'Using neural networks for classification tasks-some experiments on datasets and practical advice', *J. Opt Res.*, No. 3, 215-226
136. Hawley, D. D., Johnson, J. D. and Raina, D. (1990), 'Artificial neural systems: A new tool for financial decision-making', *Financial Analysts Journal*, Nov/Dec. 63-72
137. Hecht-Nielsen, R. (1987), 'Kolmogorov's Mapping Neural Network Existence Theorem', *Proc. IEEE 1st International Conference on Neural Networks*, June, San Diego, III-11-III-14
138. Hecht-Nielsen, R. (1991), *Neurocomputing*, Reading: Addison-Wesley
139. Henon, M. (1976), 'A Two Dimensional Mapping with a Strange Attractor', *Communications in Mathematical Physics*, 50
140. Hertz, J., Krogh, A. and Palmer, R. G. (1991), *Introduction to the Theory of Neural Computation*, Massachusetts: Addison-Wesley
141. Heskes, T. M. and Kappen, Bert (1992), 'Learning-parameter adjustment in neural networks', *Physical Review A*, A15-9a, June, 1-14
142. Hinich, M. J. and Patterson, D. M. (1985), 'Evidence of nonlinearity in daily stock returns', *Journal of Business & Economics Statistics*, 3:1, 69-77
143. Holloway, T. M. (1984), 'The economy and the federal budget: guides to the automatic effects', *Survey of Current Business* 64, July, 102-108
144. Holloway, T. M. (1989), 'Measuring the cyclical sensitivity of Federal receipts and expenditures: Simplified estimation procedures', *International Journal of Forecasting* 5, 347-360
145. Hooijmans, F. C. (1989), 'Controlekaarten voor het financieringstekort', Ministerie van Financien, Bureau Financiële Analyse en Planning, Onderzoeksnotitie 8901, Jan.
146. Hooijmans, F. C. (1992), 'Weekrapport afdeling Centraal Kasbeleid', 2 oktober, Ministerie van Financien
147. Hsieh, D. A. (1989), 'Testing for nonlinear dependence in daily foreign exchange rates', *Journal of Business*, No. 3, 1989, 339-368.
148. Hsieh, D. A. (1991), 'Chaos and nonlinear dynamics: Application to financial markets', *Journal of Finance*, XLVI:5, 1839-1877

149. Hull, J. (1989), *Options, Futures and other Derivative Securities*, Englewood Cliffs, NJ: Prentice Hall
150. Humpert, B. (1989), 'Neurocomputing in financial services', *Expert Systems for Information Management*, No. 3, 172-199
151. Jacobs, R. A. (1988), 'Increased rates of convergence through learning rate adaptation', *Neural Networks*, Vol. 1, 295-307
152. Jensen, H. L. (1992), 'Using neural networks for credit scoring', *Managerial Finance*, No. 6, 14-26
153. Jones, L. K. (1990), 'Constructive approximations for neural networks by sigmoidal functions', *Proc. of the IEEE*, No. 10, Oct. 1586-1589
154. Kamijo, K. and Tanigawa, T. (1990), 'Stock price pattern recognition: A recurrent network approach', *Proceedings of the IEEE International Joint Conference on Neural Networks*, I215-I221
155. Karels, G. V. and Prakash, A. J. (1987), 'Multivariate normality and forecasting of business bankruptcy', *Journal of Business Finance and Accounting*, 14:4, 573-593
156. Kat, H. M. (1992), 'Modeling S&P 500 futures mispricing using a neural network', Financial Management Department, UvA, 28-09-93, 1-20
157. Kayama, M., Abe, S., Takenaga, H. and Morooka, Y. (1990), 'Constructing optimal neural networks by linear regression analysis', 364-376
158. Keasey, K. and Watson, R. (1987), 'Non-financial symptoms and the prediction of small company failure: A test of Argenti's hypotheses', *Journal of Business & Accounting*, 14 (3) Autumn, 335-354
159. Keyes, J. (1990), 'Neural networks can't think, but they can learn-almost', *Computerworld*, 8 Oct.
160. Khanna, T. (1990), *Foundations of Neural Networks*, Massachusetts: Addison-Wesley
161. Kim, J. H. and Stringer, J. (eds) (1992), *Applied Chaos*, NY: John Wiley & Sons
162. Kimoto, T., Asakawa, K., Yoda, M. and Takeoda, M. (1990), 'Stock market predictions system with modular neural networks', Paper presented at the IJCNN, San Diego, 1-6
163. Klemic, G. G. (1990), 'The use of neural computing technology to develop profiles of Chapter 11 debtors who are likely to become tax delinquents'. In: Trippi, R. R. and Turban, E. (eds) (1993), *Neural Networks in Finance & Investment*, Chicago: Probus Publishing
164. Knerr, S., Personnaz, L., and Dreyfus, G. (1990), 'Single-layer revisited: a stepwise procedure for building and training a neural network'. In: Fogelman-Souillie and Herault, J. (eds) *Neurocomputing: Algorithms, Architectures and Applications*, NATO ASI Series, Springer
165. Knoop van der, H. S. (1988), 'Control charts to check yearly predictions by monthly observations', Ministry of Finance, Bureau Financial Analysis and Planning, *Research memorandum 8803*, Nov.
166. Kochan, N. (1993), 'Warsaw advances step by step', *Euromoney*, Jan. 69-72
167. Kohonen, T. (1984), *Self-organization and Associative Memory*, Springer
168. Kohonen, T. (1988), 'An introduction to neural computing', *Neural Networks*, Vol. 1, 1, 3-16
169. Kosko, B. (1992), *Neural Networks and Fuzzy Systems*, London: Prentice Hall
170. Kouam, A., Badran, F. and Thiria, S. (1992), 'Approche methodologique pour l'etude de la prevision a l'aide de reseaux de neurones', *Proceedings of Neuro-Nîmes 1992*
171. Lapedes, A. and Farber, R. (1987), 'Nonlinear signal processing using neural networks: prediction and system modelling', TR LA-UR-87-2662, Los Alamos
172. Larrain, M. (1991), 'Testing Chaos and Nonlinearities in T-Bill Rates', *Financial Analysts Journal*, Sep/Oct. 51-62
173. Le Cun, Y., Boser, B. and Denker, J. S. (1989), 'Backpropagation applied to handwritten Zip code recognition', *Neural Computation*, Vol. 1, 541-551
174. Le Cun, Y., Denker, J. S. and Solla S. A. (1990), 'Optimal brain damage'. In: Touretzky, D. S. (ed), *Advances In Neural Information Processing Systems 2*, San Mateo, California: Kaufman Publishing, 598-605
175. Le Cun, Y. (1989), 'Constrained networks for handwritten numeral recognition', Snowbird conference on 'Neural Networks for Computing', Snowbird
176. Legler, J. B. and Shapiro, P. (1968), 'The responsiveness of state tax revenue to economic growth', *National Tax Journal* XXI, 46-56
177. LeRoy, S. F. (1989), 'Efficient capital markets and martingales', *Journal of Economic Literature*, XXVII, 1583-1621
178. Levy, H. and Sarnat, M. (1970), 'International diversification of investment portfolios', *American Economic Review*, Sep. 668-692

179. Lewis, P. A. W. and Stevens, J. G. (1991), 'Nonlinear modeling of time series using multivariate adaptive regression splines (MARS)', *Journal of American Statistics Association*, 86, 864
180. Libby, R. (1975), 'Accounting ratios and the prediction of failure: Some behavioral evidence', *Journal of Accounting Research*, Spring, 150-161
181. Lin, F. C. and Lin, M. (1993), 'Analysis of financial data using neural nets', *AI Expert*, 36-41
182. Lintner, J. (1965), 'The valuation of risk assets and the selection of risky investments in stocks, portfolios and capital budgets', *Review of Economics and Statistics*, 47:1, 13-37
183. Lippman, R. P., (1987), 'An introduction to computing with neural nets', *IEEE ASSP Magazine*, April, 4-22
184. MacKay, D. J. C. (1991), 'A practical bayesian framework for backprop networks', submitted to *Neural Computation*
185. Manaster, S. and Rendleman, Jr. R. J. (1982), 'Option prices as predictors of equilibrium stock prices', *Journal of Finance*, 37, 1043-1057
186. Mandelbrot, B. B. (1971), 'When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models', *Review of Economics and Statistics*, LIII, 225-236
187. Markowitz, H. M. (1952), 'Portfolio Selection', *Journal of Finance*, 7, 77-91
188. Markowitz, H. M. (1959), *Portfolio Selection, Efficient Diversification of Investments*, NY: John Wiley
189. Maskara, A. and Noetzel, A. (1992), 'Forced simple recurrent neural networks and grammatical inference', preprint
190. McClelland, J. L. and Rumelhart, D. E. (eds) (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*, Massachusetts: MIT Press
191. McClelland, J. L. and Rumelhart, D. E. (eds) (1988), *Explorations in Parallel Distributed Processing: Computational Models of Cognition and Perceptron (software manual)*, Massachusetts: MIT Press
192. McCord Nelson, M. and Illongworth, W. T. (1991), *A Practical Guide to Neural Nets*, Massachusetts: Addison-Wesley
193. Mei, J. (1993), 'A Semiautoregression approach to the arbitrage pricing theory', *Journal of Finance*, 48, 599-620

194. Mezard, M. and Nadal, J. P. (1989), 'Learning in feedforward networks: the tiling algorithm', *Journal of Physics A*, Vol. 22, 12, 2191-2203
195. Millard, B. J. (1993), *Winning on the Stock Market*, Chichester: John Wiley & Sons
196. Minsky, M. L. and Papert, S. A. (1988), *Perceptrons: Expanded Edition*, Massachusetts: MIT Press
197. Moller, M. (1993), 'A scaled conjugate gradient algorithm for fast supervised learning', *Neural Networks*, Vol. 6, 525-533
198. Moody, J. E. (1992), 'The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems', *ANIPS* 4, 847-854
199. Moody, J. E., Levin, U. and Reh fuss, S. (1993), 'Predicting the US index of industrial production', *Neural Network World*, 3:6, 791-794
200. Müller, A. and Neumann, J. (1990), 'Classification with neural networks'. In: Bock, H. H. and Ihml, P. (eds), *Classification, Data Analysis, and Knowledge Organisation*, 32-43
201. Murtagh, F. (1990), 'A short survey of neural networks for forecasting and related problems'. In: Murtagh, F. (ed), *PASE 1990, Neural networks for statistical and economic data*, 87
202. Nazmi, N. and Leuthold, J. H. (1988), 'Forecasting economic time series that require a power transformation: case of state tax receipts', *Journal of Forecasting*, Vol. 7, 173-184
203. 'Neural networks: Theory and practice', (1989) *Byte*, Aug. 244-245
204. Nichols, N. A. (1993), 'Efficient? Chaotic? What's the new finance', *Harvard Business Review*, Mar-Apr. 50-60
205. Niranjani, Dr Mahesan (1992), 'Programming a neural network', *Neural network summer school*, Department of Engineering, University of Cambridge, 7-11 Sep. 1-6
206. Numan, H. G. K. (1990), 'Implied volatility: Niet recht maar krom', *Het Financieele Dagblad*, 1 Oct.
207. Openshaw, Stan and Wymer, C. (1990), 'A neural net classifier system for handling census data'. In: Murtagh, F. (ed), *PASE 1990, Neural networks for statistical and economic data*, 73-85
208. Ormerod, P. and Walker, T. (1990), 'Neural networks and the monetary base in Switzerland'. In: Murtagh, F. (ed), *PASE 1990, Neural networks for statistical and economic data*, 71
209. Ormerod, P. and Walker, T. (1993), 'Macroeconomic modelling of complex systems', *Neural Network World*, 3:6, 795-814

210. Panton, D. (1976), 'Chicago board call options as predictors of common stock price changes', *Journal of Econometrics*, 101–113
211. Papadourakis, G. M., Spanoudakis, G. and Gotsias, A. (1991), 'Short-term stock price forecasting using neural networks'. In: Würtz, D. and Murtagh, F. (eds) (1991), *Proceedings International Workshop on Parallel Problem Solving From Nature: Applications in Statistics and Economics PASE-91*, Zürich, 151
212. Pearlmutter, B. A. (1993), 'Fast exact multiplication by the hessian', to appear in *Neural Computation*
213. Peters, E. E. (1989), 'Fractal structure in the capital markets', *Financial Analysts Journal*, Jul/Aug. 32–37
214. Peters, E. E. (1991a), *Chaos and Order in the Capital Markets, A New View of Cycles, Prices and Market Volatility*, New York: John Wiley & Sons
215. Peters, E. E. (1991b), 'A chaotic attractor for the S&P 500', *Financial Analysts Journal*, Mar–Apr. 55–62
216. Peters, E. E. (1994), *Fractal Market Analysis*, NY: Wiley
217. Piesse, J. and Wood, D. (1992), 'Issues in assessing MDA models of corporate failure: A research note', *British Accounting Review*, Vol. 24, 33–42
218. Poprzeczko, J. (1993), 'Our expensive credit', *Polityka Financial Supplement*, No. 3, March
219. Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1992), *Numerical Recipes in C*, Cambridge: CUP
220. Priestley, M. B. (1980), 'State-dependent models: A general approach to non-linear time series analysis', *Journal of Time Series Analysis*, Vol. 1, 1
221. Priestley, M. B. (1988), *Non-linear and Non-stationary Time Series Analysis*, London: Academic Press
222. Quinlan, J. R. (1979), 'Discovering rules by induction from large classes of examples'. In: Michie, D. (ed), *Expert Systems in the Microelectronic Age*, Edinburgh: Edinburgh University Press
223. Raghupathi, W., Schkade, L. L. and Raju, B. S. (1991), 'A neural network approach to bankruptcy prediction'. In: Trippi, R. R. and Turban, E. (eds) (1993), *Neural Networks in Finance & Investment*, Chicago: Probus Publishing Company, 141–158
224. Ramanathan, Ramu (1989), *Introductory Econometrics with Applications*, San Diego: Harcourt Brace Jovanovich

225. Refenes, A. N. (ed) (1993), *NnCM'93: Proceedings of the first International Workshop on Neural Networks in the Capital Markets*, London Business School, 18–19 Nov.
226. Richardson, F. M. and Davidson, L. F. (1983), 'An exploration into bankruptcy discriminant model sensitivity', *Journal of Business Finance and Accounting*, 10:2, 195–207
227. Ripley, B. D. (1992), 'Statistical aspects of neural networks', *SemStat* 1992, 21 June, 1–70
228. Ripley, B. D. (1993a), 'Neural networks and related methods for classification', preprint
229. Ripley, B. D. (1993b), 'Statistical aspects of neural networks'. In: Barndorff-Nielsen, Jensen, J. L. and Kendall, W. S. (eds), *Networks and Chaos: Statistical and Probabilistic Aspects*, London: Chapman & Hall, 40–123
230. Ripley, B. D. (1993c), 'Neural networks and flexible regression and discrimination'. In: Mardia, K. V. (ed) *Statistics And Images*, Abingdon: Carfax, 1–24
231. Rohwer, R. (1990), 'Neural networks for time-varying data'. In: Murtagh, F. (ed), *PASE 1990*, Neural networks for statistical and economic data, 59–69
232. Ross, S. A. and Walsh, M. (1983), 'A simple approach to the pricing of risky assets with uncertain exchange rates'. In: Hawkins, R. Levich, R. and Wihlberg, C. (eds) *The Internationalization of Financial Markets and National Economic Policy*, Greenwich: JAI Press
233. Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory*, 13, 341–360
234. Rumelhart, D. E. and McClelland, J. L. (eds) (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, Massachusetts: MIT Press
235. Sanger, Terence D. (1989), 'Optimal unsupervised learning in single-layer linear feedforward neural network', *Neural Networks*, 459–473
236. Saunders, Anthony (1987), 'The inter-bank market, contagion effects and international financial crises'. In: Portes and Swoboda (eds), *Threats to International Financial Stability*, CEPR, New York: Cambridge University Press
237. Savit, R. (1992), 'Chaos on the Trading Floor'. In: Nina Hall (ed), *The New Scientist Guide to Chaos*, London: Penguin
238. Schalkoff, R. (1992), *Pattern Recognition: Statistical, Structural and Neural Approaches*, NY: John Wiley & Sons Inc

239. Scheinkman, J. A. and LeBaron, B. (1989), 'Nonlinear dynamics and stock returns', *Journal of Business*, No. 3, 311–337
240. Schiffman, W., Joost, M. and Werner, R. (1992), 'Optimization of the backpropagation algorithm for training multilayer perceptrons', the neuroprose archive, anonymous FTP; cheops.ohio-state.edu
241. Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *Journal of Finance*, 19, 425–442
242. Shkurti, W. J. and Winefordner, D. (1989), 'The politics of state revenue forecasting in Ohio, 1984–1987: A case study and research implications', *International Journal of Forecasting* 5, 361–371
243. Simon, H. (1982), *Models of Bounded Rationality*, Cambridge: MIT Press
244. Sirat, J. A. and Nadal, J. P. (1990), 'Neural trees: a new tool for classification', *Network*, Vol. 1, 423–438
245. Smith, C. J. (1991) 'A neural network — could it work for you', *Financial Executive*, May/June, 26–30
246. Smith, M. (1991), 'Neural networks: Do they compute?', *Best's Review*, 70–74
247. Smith, M. (1993), *Neural Networks for Statistical Modelling*, NY: VNR
248. Speidel, L. S. and Sappenfield, R. (1992), 'Global diversification in a shrinking world', *Journal of Portfolio Management*, 19, 57–67
249. Stephan, J. A. and Whaley, R. E. (1990), 'Intraday price change and trading volume relations in the stock and stock options market', *Journal of Finance*, 55, 191–220
250. Stein, R. (1993a), 'Preprocessing data for neural networks', *AI Expert*, 32–37
251. Stein, R. (1993b), 'Selecting data for neural networks', *AI Expert*, 42–47
252. Sterge, A. J. (1989), 'On the distribution of financial futures price changes', *Financial Analyst Journal*, May/June
253. Still, D. B. (1991), 'A philosophical foundation for credit analysis', *The Journal of Commercial Bank Lending*, Nov. 43–47
254. Sugihara, G. and May, R. M. (1990), 'Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series', *Nature*, 19 Apr. 734–741
255. Surkan, A. and Singleton, J. (1991), 'Neural networks for bond rating improved by multiple hidden layers'. In: *Proceedings of the IEEE International Conference on Neural Networks*, Jul. II 157–162
256. Taffler, R. J. (1982). 'Forecasting company failure in the UK using discriminant analysis and financial ratio data', *Journal of the Royal Statistical Society, Series A*, Vol. 145, Part 3, 342–358
257. Taffler, R. J. (1984), 'Empirical models for the monitoring of UK corporations', *Journal of Banking and Finance*, 199–227
258. Taffler, R. J. and Tseung, M. (1984), 'The audit going concern in practice', *The Accounting Magazine*, July
259. Tam, K. Y. and Kiang, M. (1990), 'Predicting bank failures: a neural network approach', *Applied Artificial Intelligence; an International Journal*, Vol. 4, No. 4, 265–282
260. Thompson, J. M. T. and Stewart, H. B. (1991), *Nonlinear Dynamics and Chaos*, NY: John Wiley & Sons
261. Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis*, Lectures Notes Statistics, Vol. 21, Springer
262. Toussaint, G. T. (1974), 'Bibliography on estimation of misclassification', *IEEE Transactions on Information Theory*, Vol. 20, 4, 472–479
263. Treleaven, P. and Goonatilake, S. (1991), 'Intelligent financial technologies'. In: Würtz, D. and Murtagh, F. (eds), *Proceedings International Workshop on Parallel Problem Solving From Nature: Applications in Statistics and Economics PASE-91*, Zürich, 7–26
264. Trigueros, D. and Berry, R. H. (1991), 'Applying neural networks to the extraction of knowledge from accounting reports: A classification study', School of Information Systems, University of East Anglia, Norwich, NR4 7TJ, Mar. 1–28
265. Trippi, R. R. and DeSieno, D. (1992), 'Trading equity index futures with neural network', *The Journal of Portfolio Management*, Fall, 27–33
266. Trippi, R. R. and Turban, E. (eds) (1993), *Neural Networks in Finance & Investment*, Chicago: Probus Publishing Company
267. Trzcinka, C. (1986), 'On the number of factors in the arbitrage pricing model', *Journal of Finance*, 41, 347–368
268. Tucker, A. L., Madura, J. and Chiang, T. C. (1991), *International Financial Markets West*, St Paul, 167
269. Unnikrishnan, K. P. and Venugopal, K. P. (1993), 'Alopec: a correlation based learning algorithm for feed-forward and recurrent neural networks', Technical Report GMR-7919
270. Utans, J. and Moody, J. (1991), 'Selecting neural network architectures via the prediction risk: Application to corporate bond

- rating prediction', *Proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street*, Los Alamitos
271. Varfis, A. and Versino, C. (1990), 'Univariate economic time series forecasting by connectionist methods', *Proc. International Neural Network Conference*, July, Paris
 272. Varfis, A. and Versino, C. (1990), 'Neural networks for economic time series forecasting'. In: Murtagh, F. (ed), *PASE 1990*, Neural networks for statistical and economic data, 155–159
 273. Wasserman, P. D. (1989), *Neural Computing: Theory and Practice*, NY: VNR
 274. Weers, D. (1991), 'Neural network based forecasting in financial market place', in Würtz, D. and Murtagh, F. (eds), *Proceedings International Workshop on Parallel Problem Solving From Nature: Applications in Statistics and Economics PASE-91*, Zürich, 55
 275. Weigend, A. S., Huberman, B. A. and Rumelhart, D. E. (1990), 'Predicting the future: A connectionist approach', *International Journal of Neural Systems*, Vol. 1, 3, 193–209
 276. Weigend, A. S. and Gershenfeld, N. A. (1993), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Massachusetts: Addison-Wesley
 277. Weiss, S. M. and Kapouleas, I. (1989), 'An empirical comparison of pattern recognition, neural nets and machine learning classification methods', *Proceedings 11th IJCAI*, Detroit, 781–787
 278. Werbos, P. J. (1990), 'Backpropagation through time: what it does and how to do it', *Proc. IEEE*, No. 10, Oct. 1550–1560
 279. White, H. (1988), 'Economic prediction using neural networks: The case of IBM daily stock returns'. In: Trippi, R. R. and Turban, E. (eds) (1993), *Neural Networks in Finance and Investing*, Chicago, Probus Publishing Company, 315–329
 280. White, H. (1992), *Artificial Neural Networks: Approximation & Learning Theory*, Oxford: Blackwell
 281. Widrow, B. and Hoff, M. E. (1960), 'Adaptive switching circuits', IRE WESCON Convention Record, NY, 96–104
 282. Widrow, B. (1990), '30 years of adaptive neural networks: perceptron, madaline, and backpropagation', *Proc. IEEE*, No. 10, Oct. 1415–1442
 283. Wilcox, J. W. (1992), 'Taming frontier markets', *Journal of Portfolio Management*, 19, 51–55
 284. Windsor, C. G. and Harker, A. H. (1988), 'Multi-variate financial index prediction: A neural network study', *nEuro 88 Conference*, 357
 285. Wong, F. S., Wang, P. Z., Goh, T. H. and Quek, B. K. (1992), 'Fuzzy neural systems for stock selection', *Financial Analysts Journal*, Jan-Feb. 47–74
 286. Wood, D. and Piesse, J. (1987), 'Higgledy piggledy bankruptcy', Manchester Business School Working Paper Series, No. 148
 287. Wood, D. and Piesse, J. (1988), 'The information value of failure predictions in credit assessment', *Journal of Banking and Finance*, No. 12, 275–292
 288. Würtz, D. and Murtagh, F. (eds) (1991), *Proceedings International Workshop on Parallel Problem Solving From Nature: Applications in Statistics and Economics PASE-91*, Zürich, obtainable from Eurostat Statistical Office of the European Communities, Luxembourg
 289. Würtz, D. and De Groot, C. (1991), 'Analysis of univariate time series with connectionist nets: a case study of two classical examples', *Neurocomputing*, 1991, 177–192
 290. Würtz, D. and De Groot, C. (1992a), 'Forecasting time series with connectionist nets: Applications in statistics, signal processing and economics'. In: Belli, F. and Radermacher, F. J. (eds), *Lecture Notes in Artificial Intelligence*, 604, Heidelberg: Springer
 291. Würtz, D. and De Groot, C. (1992b), 'ZIP-Zürich information processing: Predicting currency exchange rates using antithetic connectionist function approximators', Working paper, Interdisziplinäres Projektzentrum für Supercomputing, ETH-Zentrum and Institut für Theoretische Physik, ETH-Honggerberg, Zürich
 292. Würtz, D. and De Groot, C. (1992c), 'Nonlinear time series analysis with connectionist nets: Towards a robust methodology', *Proceedings of the SPIE Conference on Applications of Artificial Neural Networks*, Orlando
 293. Würtz, D., De Groot, C., Schütterle, B., Sala, W., Wenger, D. and Ünsel, S. (1993), 'Analysis and prediction of currency exchange rates', CeBIT Hannover, 1–3
 294. Yoon, Y. and Swales, G. (1991), 'Predicting stock price performance: A neural network approach', *Proceedings of the 24th Annual Hawaii International Conference on System Science*, 156–162
 295. Yule, G. U. (1927), 'On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers', *Phil. Trans. Royal Society London A226*, 267
 296. Zeeman, E. C. (1974), 'On the unstable behavior of stock exchanges', *Journal of Mathematical Economics*, 39–49

Предметный указатель

AIBOR (Amsterdam one-month inter-bank offered rate) 82, 108

APT (arbitrage pricing theory), см. также теория арбитражного ценообразования 97, 120

ARCH (auto regressive conditional heteroschedasticity), см. также авторегрессионная условная гетероскедастичность 76, 137, 194

ARIMA 50, 52, 64, 73, 78, 137

ARMA 39

ARV (average relative variance) см. также средняя относительная дисперсия 47, 87

BIC (Bayesian information criterion), см. также байесовский информационный критерий 84

CAPM (capital asset pricing model), см. также модель ценообразования на рынке капиталовложений 97, 119, 146

CMA (crossing-moving averages), см. также перекрестное правило скользящих средних 197

CPN (counter propagation network), см. также нейронная сеть со встречным распространением 24

EMH (efficient market hypotheses), см. также эффективного рынка гипотеза 56, 118, 193

EOE (European option exchange), см. также Европейская биржа опционов 96

Excel, лист электронной таблицы (Excel worksheet) 53, 127

FKSEC-модель 79

FPE (final prediction error), см. также итоговая ошибка прогноза 174

GARCH-модель 194

IBM viii, 195

LDA (linear discriminant analysis), см. также дискриминантный анализ 29

logit, см. также логистическая регрессия 151, 153

LPM (linear probability model), см. также линейная вероятностная модель 153

MA-правило (moving-average rule), см. также правило скользящего среднего 194

MARS (multivariate adaptive regression splines), см. также многомерные адаптивные регрессионные сплайны 49

MBPN (multilayer backpropagation network), см. также нейронная сеть многослойная с алгоритмом обратного распространения ошибки 64, 69, 162

MDA (multiple discriminant analysis), см. также множественный дискриминантный анализ 151, 153, 184

MLP (multilayer perceptron), см. также нейронная сеть с прямой связью и многослойный перцептрон 26

MSCI (Morgan Stanley capital international index), см. также международный индекс капитализации Моргана Стэнли 141

MSE (mean squared error), см. также средняя квадратичная ошибка 46

NBIC (normalised Bayesian information criterion), см. также нормализованный байесовский информационный критерий 84

NIB (Dutch Investment Bank), см. также Голландский инвестиционный банк 168

QDA (quadratic discriminant analysis), см. также дискриминантный анализ 29

RMSE (square root of the mean squared error), см. также квадратный корень из средней квадратичной ошибки 46, 67, 124

SME (small and medium-sized enterprises), см. также малые и средние предприятия и МСП 154

STAR (smooth threshold autoregressive model), см. также гладкая пороговая авторегрессионная модель 40, 41

TAR (threshold autoregressive model), см. также пороговая авторегрессионная модель 39, 41, 49, 50, 76, 137

Theils μ 51

TRB-правило (trading-range break rule), см. также правило превышения предела изменения цен 194

VAR, метод 137

Z-модель Альтмана (Altman's 'Z' score model) 160, 183

Атокорреляция (autocorrelation) 63
авторегрессионная условная гетероскедастичность, см. также ARCH 76, 137, 194

Акаике (Akaike) 47

аксон (axon) 2

активации функция (activation function), см. также функция активации 3

— сигмоидная (sigmoid) 4

активация (activation)

—, уровень ($\sim$ level) 90

Альтман (Altman), см. также Z-модель Альтмана 160, 183

Амстердамский индекс курсов акций (Amsterdam general stock index) 83

анализ главных компонент (princals) 160, 171

априорные знания (a priori knowledge) 7

аттрактор (attractor)

— предельный цикл (limit cycle $\sim$) 61

— странный (strange $\sim$) 61, 69

— точечный (point $\sim$) 61

Бабушкино кодирование (grand-mother coding) 31

байесовский информационный критерий, см. также BIC 84

банкротство корпораций (corporate failure) 183

безусловное ожидание (unconditional expectation) 127

белый шум (white noise) 124

ближайший сосед (nearest neighbour) 29

Блэк (Black), см. также модель Блэка-Шоулса 98

Бокс (Box), см. также метод Бокса-Дженкинса 50, 51, 64, 73

буквенная переменная (nominal variable) 160

Вероятностная функция плотности (probability density function) 29
вероятностный (probabilistic) нейрон 3

веса (weights), см. также синаптические коэффициенты 3

вектор (weight vector) 23

взбалтывание (scrambling) 12

Вигенд (Weigend), см. также Вигенда отношение 48, 75, 128

Вигенда отношение (Weigend ratio) 48, 75, 128

возбуждающая связь (excitatory link) 2

волатильность подразумеваемая (implied volatility) 106

временная цена опциона (option time value) 104

время до исполнения (time to maturity) 106

Гамма опциона (option gamma) 110
генетический алгоритм (genetic algorithm) 15

гиперплоскость (hyperplane) 27

гладкая пороговая авторегрессионная модель, см. также STAR 40, 41

Голландский инвестиционный банк, см. также NIB 168

градиентный шаг, величина (gradient step size), см. также коэффициент обучения 9

Данные за один день (intraday data) 99

Дарбина-Уотсона показатель (Durbin-Watson statistic) 86

движущееся окно (moving window) 114

дельта опциона (option delta) 109

дельта-правило обобщенное (generalised delta rule) 10

дельта-хеджирование (delta hedging) 98

детерминированный (deterministic) нейрон 3

Дженкинс (Jenkins), см. также метод Бокса-Дженкинса 50, 51, 64, 73

диверсификация (diversification) 140

динамическая сеть (dynamic network) 21

динамический (dynamic) нейрон 3

дискриминантный анализ (discriminant analysis)

— квадратичный (quadratic ~) 29

— линейный (linear ~) 29

доводка (postprocessing) 42, 48

Евклидово расстояние (Euclidean distance) 91

Европейская биржа опционов (Амстердам), см. также EOE 96

ежедневные данные (interday data) 99

Замораживание (annealing) 15

Иерархическая классификация (hierarchical clustering) 91

импульс (momentum) 13

инвариантность преобразования (translation invariance) 201

информационный критерий (information criterion) 47

искусственный интеллект (artificial intelligence) 1

историческая волатильность (historical volatility) 107

итоговая ошибка прогноза, см. также FPE 174

Календарный эффект (calendar effect) 81

квадратичный дискриминантный анализ, см. также QDA 29

квадратный корень из средней квадратичной ошибки, см. также RMSE 46, 67, 124

классификация (classification) 1, 7, 26

кодирование выхода (output coding) 31

Кокрана-Оркютта итерационный процесс (Cochrane-Orcutt iterative procedure) 86

Колмогоров 84

комбинированная гипотеза (joint hypothesis) 193

коэффициент обучения (learning rate) 9

кредитный риск (credit risk) 151

критерий максимума правдоподобия (maximum likelihood) 9

критерий ошибки (fitness) 15

крутизна (steepness) сигмоида 4

Ликвидность (liquidity) 98

линейная вероятностная модель, см. также LPM 153

линейная отделимость (linear separation) 27

линейный дискриминантный анализ, см. также LDA 29

логистическая регрессия (logistic regression), см. также logit 153

логистическое отображение (logistic map) 59

локальный метод оптимизации (local optimization method) 14

локальный минимум (local minimum) 12, 124

лямбда опциона (option lambda) 110

M1, см. также предложение денег 83

максимума правдоподобия критерий (maximum likelihood) 9

малые и средние предприятия, см. также МСП и SME 154

матрица ошибок (confusion matrix) 161

международный индекс капитализации Моргана Стэнли, см. также MSCI 141

метод k ближайших соседей (k -nearest-neighbours approach) 29, 168, 172

метод Бокса-Дженкинса, см. также ARIMA 50, 51, 64, 73

Министерство финансов Голландии (Dutch Ministry of Finance) 78

многомерные адаптивные регрессионные сплайны, см. также MARS 49

многослойный перцептрон (multilayer perceptron) 26

множественный дискриминантный анализ, см. также MDA 151, 153, 184

модель Блэка-Шоулса (Black-Scholes model) 98

модель ценообразования на рынке капиталовложений, см. также CAPM 97

МСП, см. также малые и средние предприятия и SME 154

мультиколлинеарность (multicollinearity) 112

мутация (mutation) 15

Начальные веса, выбор (weight initialisation) 11

нейрон (neuron) 2

— вероятностный 3

— детерминированный 3

— динамический 3

— скрытый 5

— статический 3

нейронная сеть (neural network)

— адаптивная (adaptive ~) 116

— архитектура (~architecture) 33

— динамическая (dynamic ~) 21

— с временной задержкой (time delay ~) 21

— со встречным распространением (counter propagation ~), см. также CPN 24

— Кохонена (Kohonen ~), см. также ~самоорганизующаяся 22

— многослойная с алгоритмом обратного распространения ошибки (multilayer backpropagation ~), см. также MBPN 64, 69, 162

—, объем (~size) 31

— с прямой связью (feed-forward ~), см. также многослойный перцептрон 5, 26

— рекуррентная (recurrent ~) 5

— Хопфилда (Hopfield ~) 22

— самоорганизующаяся (self-organising ~), см. также ~Кохонена 22

—, топология (topology) 5

нелинейная отделимость (non-linear separation) 27

нелинейная фильтрация (non-linear filtering) 22

нелинейные динамические процессы (non-linear dynamics) 1, 56, 62

нелинейный анализ главных компонент (non-linear principal components analysis) 151

непараметрический метод (non-parametric method) 29

неравносность (overshooting) 98

нервная клетка (nerve cell) 2

неувязки в бюджете (asset-liability mismatch) 79

неэффективность рынка (market inefficiency) 205

нормализованный байесовский информационный критерий, см. также NBIC 84

Обменный курс (currency exchange) 45

обобщение (generalisation) 28, 164

обобщенное дельта-правило (generalised delta rule) 10

обработка пакетная (batching) 12

обратное распространение ошибки (backpropagation)

—, алгоритм (~algorithm) 10

— многослойный (multilayer ~) 10

— стохастический 11

обратное распространение ошибки, алгоритм (продолжение)

— «эпохальный» (per-epoch) 13

обращение тренда (trend reversion) 58

обучающее множество (training set) 7

обучающий алгоритм (learning algorithm) 8

обучение (learning)
 —, алгоритм 8
 — контролируемое (supervised ~) 7
 —, коэффициент (learning rate) 9
 — самостоятельное (unsupervised ~) 22
 —, сходимость (convergence in learning) 8
 ограниченной рациональности (bounded rationality) принцип 56
 ограничивающий параметр (decay term) 204
 опцион колл (call option) 97
 открытый интерес (open interest) 105
 отростки (dendrites) 2
 ошибка 1-го рода (type I error) 161, 185
 ошибка 2-го рода (type II error) 161, 185
 ошибка классификации (misclassification) 34, 161, 183
 ошибка средняя квадратичная (mean squared error) 9, 46
 Перекрестное подтверждение (cross-validation) 18, 151, 163
 — с одним в остатке (leave-one-out) 19
 перекрестное правило скользящих средних, см. также СМА 197
 переменная (variable)
 —, взаимодействие (interaction) 90
 —, вклад (contribution) 94, 128, 166
 —, двоичная (binary) 189
 — непрерывная (continuous) 189
 — описательная (explanatory) 101
 —, распознающая состояние (state recognition) 101
 переобучение (overfitting), см. также эффект бабушкиного воспитания 16
 погода (weather) 83
 подсказка (hint) 209
 подтверждающее множество (validation set) 18
 Польша 151

порог (bias) 3
 пороговая авторегрессионная модель, см. также TAR 39, 41, 49, 50, 76, 137
 портфельные инвестиции (portfolio investment) 138
 порядковая переменная (ordinal variable) 160
 последовательное уменьшение (concurrent descent) ошибки 18
 потенциал (potential) нейрона 3
 правила расчета Primacheque (Primacheque regulation) 78
 правило превышения предела изменения цен, см. также TRB-правило 194
 правило скользящего среднего, см. также МА-правило 194
 предварительная обработка (preprocessing) 27, 42, 111, 144
 предложение денег (money supply), см. также M1 83
 представительность (representativity) данных 19
 премия за риск (risk premium) 97
 прогноз, не зависящий от модели (model free prediction) 76
 пропущенные значения (missing values) 80
 пространство входов (input space) 27
 пространство задержки (lag space) 38
 пут-колл соотношение (put call ratio) 109

Развивающиеся рынки (emerging markets) 139
 разница цен спроса и предложения (bid-ask spread) 106
 разрешающая способность (decisiveness), см. также решающее классификационное правило 34, 131
 распознавание образов (pattern recognition) 1
 расстояние Махаланобиса (Mahalanobis distance) 168
 расширенное входное пространство (augmented space) 3
 регуляризация (regularisation) 19, 204

решающее классификационное правило (decisiveness-of-classification criterion) 131
 Саморегуляция (self regulation) 62, 64
 сезонность (seasonal pattern) 81
 сеть с радиальными функциями (radial basis network) 209
 сигмоид (sigmoid) 4
 синапс (synaps) 2
 — с запаздыванием (delayed ~) 4
 синаптические коэффициенты (synaptic coefficients), см. также веса 3
 систематическая ошибка при оптимизации (optimisation bias) 197
 складного ножа (jack-knife) метод 19, 163
 скрещивание (cross-over) 15
 скрытый (hidden) нейрон 5
 скрытый слой (hidden layer) 5
 сложные отображения (complex mapping) 2
 случайное блуждание (random walk) 56
 справедливость (conscience) 23
 средняя квадратичная ошибка (mean squared error), см. также MSE 9, 46
 средняя относительная дисперсия, см. также ARV 47, 87
 статический (static) нейрон 3
 стохастический градиентный (stochastic gradient) метод 9
 Теорема Колмогорова о существовании отображения (Kolmogorov's mapping existence theorem) 84
 теория арбитражного ценообразования (arbitrage pricing theory), см. также АРТ 97, 120
 теория хаоса (chaos theory) 56
 тест на допустимые отклонения (tolerance test) 112
 технический анализ (technical analysis) 193
 тормозящая связь (inhibitory link) 2
 Уменьшение весов (weight-decay) 19

Фаза цикла деловой активности (business cycle) 83
 фазовое пространство (phase space) 38
 Филипс (Philips) 96
 функция активации (activation function) 3
 — линейная (linear ~) 4
 — с насыщением (saturated) 4
 — многопороговая (multithreshold ~) 4
 — сигмоидная (sigmoid) 4
 —, крутизна (steepness) 4
 — ступенчатая (step-function) 4
 функция стоимости (cost function) 19

Хенон (Henon) 68
 Холта-Уинтерса (Holt Winters) метод 51

Цена исполнения (exercise price) 102
 Центральное плановое управление Голландии (Dutch Central Planning Office) 79

Чартист (chartist) 194

Шоулс (Scholes), см. также модель Блэка-Шоулса 98
 штрафное слагаемое (penalty) 19, 204
 шум (noise) 1, 16, 98

Эпоха (epoch) 9
 эффект бабушкиного воспитания (grandmother effect), см. также переобучение 16
 эффективного рынка гипотеза, см. также ЕМН 56, 97

Юнилевер (Unilever) 193

Ядро (nucleus) 2